United States Government Accountability Office

GAO

Report to Congressional Requesters

October 2009

WARFIGHTER SUPPORT

Independent Expert Assessment of Army Body Armor Test Results and Procedures Needed Before Fielding

GAO-10-119

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Warfighter Support. Independent Expert Assessment of Army Body Armor Test Results and Procedures Needed Before Fielding	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Government Accountability Office,441 G Street NW,Washington,DC,20548**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**110**	

October 2009

WARFIGHTER SUPPORT

Independent Expert Assessment of Army Body Armor Test Results and Procedures Needed Before Fielding

GAO Accountability•Integrity•Reliability

Highlights

Highlights of GAO-10-119, a report to congressional requesters

Why GAO Did This Study

The Army has issued soldiers in Iraq and Afghanistan personal body armor, comprising an outer protective vest and ceramic plate inserts. GAO observed Preliminary Design Model testing of new plate designs, which resulted in the Army's awarding contracts in September 2008 valued at a total of over $8 billion to vendors of the designs that passed that testing. Between November and December 2008, the Army conducted further testing, called First Article Testing, on these designs. GAO is reporting on the degree to which the Army followed its established testing protocols during these two tests. GAO did not provide an expert ballistics evaluation of the results of testing. GAO, using a structured, GAO-developed data collection instrument, observed both tests at the Army's Aberdeen Test Center, analyzed data, and interviewed agency and industry officials to evaluate observed deviations from testing protocols. However, independent ballistics testing expertise is needed to determine the full effect of these deviations.

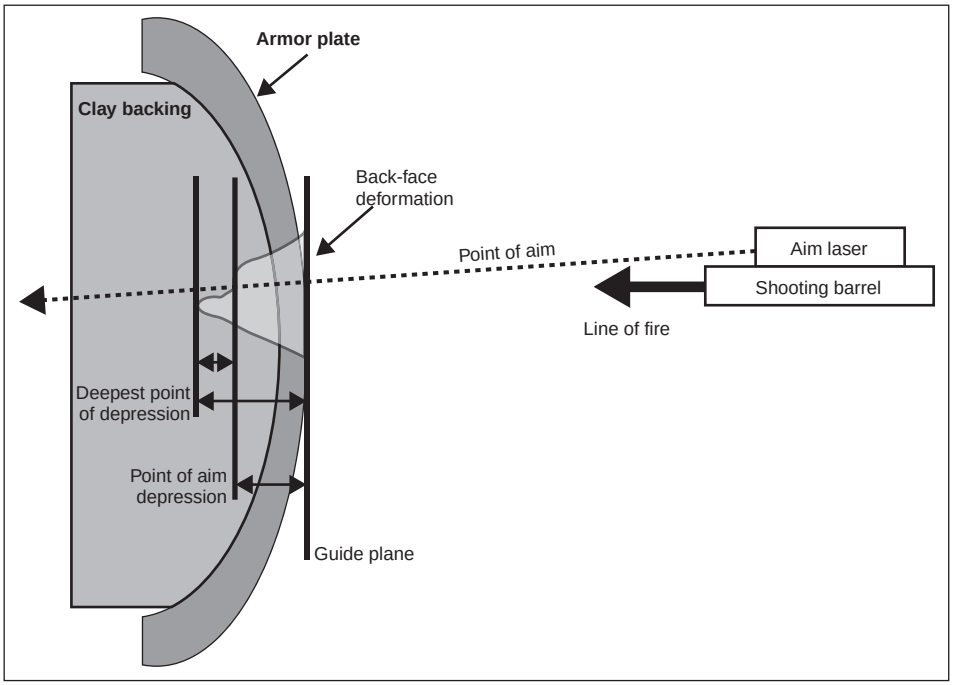
What GAO Recommends

GAO makes several recommendations, which are discussed on the next page, including to provide for an independent assessment of First Article Testing data, to assess the need to change Army's procedures based on that assessment, documenting this and all other key decisions made, and to provide for an external peer review of Aberdeen Test Center's protocols, facilities, and instrumentation.

View GAO-10-119 or key components.
For more information, contact William M. Solis at (202) 512-8365 or solisw@gao.gov.

What GAO Found

During Preliminary Design Model testing the Army took significant steps to run a controlled test and maintain consistency throughout the process, but the Army did not always follow established testing protocols and, as a result, did not achieve its intended test objective of determining as a basis for awarding contracts which designs met performance requirements. In the most consequential of the Army's deviations from testing protocols, the Army testers incorrectly measured the amount of force absorbed by the plate designs by measuring back-face deformation in the clay backing at the point of aim rather than at the deepest point of depression. The graphic below depicts the difference between the point of aim and the deepest point.



Source: GAO analysis.

Army testers recognized the error after completing about a third of the test and then changed the test plan to call for measuring at the point of aim and likewise issued a modification to the contract solicitation. At least two of the eight designs that passed Preliminary Design Model testing and were awarded contracts would have failed if measurements had been made to the deepest point of depression. The deviations from the testing protocols were the result of Aberdeen Test Center's incorrectly interpreting the testing protocols. In all these cases of deviations from the testing protocols, the Aberdeen Test Center's implemented procedures were not reviewed or approved by the Army and Department of Defense officials responsible for approving the testing protocols. After concerns were raised regarding the Preliminary Design Model testing, the decision was made not to field any of the plate designs awarded contracts until after First Article Testing was conducted.

During First Article Testing, the Army addressed some of the problems identified during Preliminary Design Model testing, but GAO observed instances in which Army testers did not follow the established testing protocols and did not maintain internal controls over the integrity and reliability of data, raising questions as to whether the Army met its First Article Test objective of determining whether each of the contracted designs met performance requirements. The following are examples of deviations from testing protocols and other issues that GAO observed:

- The clay backing placed behind the plates during ballistics testing was not always calibrated in accordance with testing protocols and was exposed to rain on one day, potentially impacting test results.

- Testers improperly rounded down back-face deformation measurements, which is not authorized in the established testing protocols and which resulted in two designs passing First Article Testing that otherwise would have failed. Army officials said rounding is a common practice; however, one private test facility that rounds told GAO that they round up, not down.

- Testers used a new instrument to measure back-face deformation without adequately certifying that the instrument could function correctly and in conformance with established testing protocols. The impact of this issue on test results is uncertain, but it could call into question the reliability and accuracy of the measurements.

- Testers deviated from the established testing protocols in one instance by improperly scoring a complete penetration as a partial penetration. As a result, one design passed First Article Testing that would have otherwise failed.

With respect to internal control issues, the Army did not consistently maintain adequate internal controls to ensure the integrity and reliability of test data. In one example, during ballistic testing, data were lost, and testing had to be repeated because an official accidentally pressed the delete button and software controls were not in place to protect the integrity of test data. Army officials acknowledged that before GAO's review they were unaware of the specific internal control problems we identified.

As a result of the deviations from testing protocols that GAO observed, four of the five designs that passed First Article Testing and were certified by the Army as ready for full production would have instead failed testing at some point during the process, either during the Preliminary Design Model testing or the subsequent First Article Test. Thus, the overall reliability and repeatability of the test results are uncertain. Although designs passed testing that would not have if the testing protocols were followed, independent ballistics experts have not assessed the impact of the deviations from the testing protocols to determine if the effect of the deviations is sufficient to call into question the ability of those designs to meet requirements. Vendors whose designs passed First Article Testing have begun production of plates. The Army has ordered 2,500 sets of plates (at two plates per set) from these vendors to be used for additional ballistics testing and 120,000 sets of plates to be put into inventory to address future requirements. However, to date, none of these designs have been fielded because, according to Army officials, there are adequate numbers of armor plates produced under prior contracts already in the inventory to meet current requirements.

GAO's Recommendations

To determine what effect, if any, the problems GAO observed had on the test data and on the outcomes of First Article Testing, the Army should provide for an independent ballistics evaluation of the First Article Testing results by ballistics and statistical experts external to the Department of Defense before any armor is fielded to soldiers under this contract solicitation. Because DOD did not concur with this recommendation, GAO added a matter for congressional consideration to this report suggesting that Congress direct DOD to either conduct such an independent external review of these test results or repeat First Article Testing.

To better align actual test practices with established testing protocols during future body armor testing, the Army should assess the need to change its test procedures based on the outcome of the independent experts' review and document these and all other key decisions made to clarify or change the testing protocols during future body armor testing. Although DOD did not agree that an independent expert review of test results was needed, DOD stated it will address protocol discrepancies identified by GAO as it develops standardized testing protocols. DOD also agreed to document all decisions made to clarify or change testing protocols.

To improve internal controls over the integrity and reliability of test data for future testing as well as provide for consistent test conditions and comparable data among tests, the Army should provide for an independent external peer review of Aberdeen Test Center's body armor testing protocols, facilities, and instrumentation to ensure that proper internal controls and sound management practices are in place. DOD generally concurred with this recommendation, but stated that it will also include DOD members on the review team.

Contents

Abbreviations

ASTM	American Society for Testing and Materials
DOD	Department of Defense
DODIG	Department of Defense Inspector General
ESAPI	Enhanced Small Arms Protective Insert
FSAPV-E	Flexible Small Arms Protective Vest-Enhanced
FSAPV-X	Flexible Small Arms Protective Vest-X level
MDA	Milestone Decision Authority
NIJ	National Institute of Justice
PEO	Program Executive Office
XSAPI	Small Arms Protective Insert-X level

GAO
Accountability * Integrity * Reliability

United States Government Accountability Office
Washington, DC 20548

October 16, 2009

Congressional Requesters

Since combat operations began in Afghanistan after September 11, 2001, and in Iraq in 2003, U.S. forces have been subjected to frequent and deadly attacks from insurgents using improvised explosive devices, mortars, rocket launchers, and increasingly lethal ballistic threats. To protect the military and civilian personnel of the Department of Defense (DOD) against these ballistic threats, since 2003 the U.S. Central Command has required that DOD personnel in its area of operations be issued the Interceptor Body Armor system, comprising ceramic plates that are inserted into pockets of an outer protective vest. Over the past several years, the media and Congress have raised concerns about whether the Army has adequately evaluated and tested this body armor solution and about the transparency of the Army's body armor testing. Additionally, several audits have found problems with the Army's body armor testing programs. For example, in 2009, both the DOD Inspector General and Army Audit Agency reported that the Army had not followed established test procedures during prior tests of body armor plates.[1]

In 2007, we reported to the House and Senate Armed Services Committees and testified to the House Armed Services Committee about the Army's and Marine Corps's individual body armor systems.[2] In that report we found that the Army relied on several controls to ensure that body armor met performance requirements, including testing at National Institute of Justice (NIJ)-certified testing facilities. Later, under the Comptroller General's authority, we observed the testing of body armor solutions submitted under a May 2007 Army contract solicitation for four categories of body armor—specifically, the Enhanced Small Arms Protective Insert (ESAPI), the Small Arms Protective Insert-X level (XSAPI), the Flexible Small Arms Protective Vest-Enhanced (FSAPV-E), and the Flexible Small Arms Protective Vest-X level (FSAPV-X). While present, we observed the

[1] DOD Inspector General, *DOD Testing Requirements for Body Armor*, D-2009-047 (Arlington, Va.: Jan. 29, 2009); and U.S. Army Audit Agency, *Body Armor Testing: Program Executive Office, Soldier*, A-2009-0086-ALA (Alexandria, Va.: Mar. 30, 2009).

[2] GAO, *Defense Logistics: Army and Marine Corps' Individual Body Armor System Issues*, GAO-07-662R (Washington, D.C.: Apr. 26, 2007); and *Defense Logistics: Army and Marine Corps' Body Armor Requirements, Controls, and Other Issues*, GAO-07-911T (Washington, D.C.: June 6, 2007).

test procedures utilized by Army testers, spoke with Army officials, and compared our observations with established testing protocols. The purchase descriptions accompanying the contract solicitation announcement identified the test procedures to be followed during the first round of testing—called Preliminary Design Model testing. Traditionally, Army body armor testing had been performed at an NIJ-certified facility. However, one manufacturer of flexible small arms protective vests, which had failed previous testing conducted for the Program Executive Office (PEO) Soldier at an NIJ-certified facility, made allegations that the PEO Soldier and the facility had wrongly failed its designs. As a result of these allegations, the Army decided instead to conduct testing for this current solicitation at the Army's Aberdeen Test Center, which had not performed testing of Interceptor Body Armor for PEO Soldier since the 1990s. Additionally, PEO Soldier decided not to provide any on-site testing oversight to avoid any appearance of bias against that manufacturer.[3]

Preliminary Design Model testing was conducted by the Army's Aberdeen Test Center from February 2008 though June 2008. The objective of the Preliminary Design Model testing was to determine whether candidate designs submitted under the solicitation met required ballistics performance specifications and would be awarded a production contract.[4] In October 2008, on the basis of the Preliminary Design Model testing results, the Army awarded four 5-year indefinite delivery/indefinite quantity[5] contracts at a total of over $8 billion for the production of the ESAPI and the XSAPI—two categories of ceramic plates. No FSAPV-E or FSAPV-X solutions passed the testing. The Army decided to repeat testing, through First Article Testing, of all of the ESAPI and XSAPI plates that

[3] The designs submitted by that manufacturer also failed Preliminary Design Model testing at Aberdeen Test Center.

[4] The armor plate contracts require First Article Testing, in accordance with the Federal Acquisition Regulation, Subpart 9.3, to ensure the contractor can furnish a product that conforms to all contract requirements for acceptance. However, the standard Federal Acquisition Regulation First Article Testing clause allows the government to waive First Article Testing if a design has already been demonstrated to meet the required specifications.

[5] Indefinite delivery/indefinite quantity contracts provide for an indefinite quantity of supplies or services during a fixed period of time. These types of contracts are generally used when agencies are unable to predetermine, above a specified minimum, the precise quantities of supplies or services that the government will require during the contract period.

were awarded production contracts to determine whether these plate designs indeed met the required ballistics performance specifications before fielding the plates. The Aberdeen Test Center conducted First Article Testing between November 2008 and December 2008.

In connection with the Army's decision to conduct First Article Testing on each of the designs that passed Preliminary Design Model testing and that were awarded contracts, the House Armed Services Committee and its Subcommittee on Air and Land Forces requested that we observe this follow-on First Article Testing to assess the degree to which testing was conducted according to the established testing protocols. After completing our analysis of both the Preliminary Design Model testing and First Article Testing of body armor solutions, we are reporting on the degree to which the Army followed its established testing protocols during (1) Preliminary Design Model testing of the ESAPI, XSAPI, FSAPV-E and FSAPV-X and (2) First Article Testing of the ESAPI and XSAPI models that were awarded contracts after the Preliminary Design Model testing.[6] We did not provide an expert ballistics evaluation of the results of testing.

To conduct our review, we observed Preliminary Design Model testing and First Article Testing at the Army's Aberdeen Test Center in Aberdeen, Maryland. We observed testing from inside the video viewing room and firing lanes and also from the conditioning, X-ray, and physical characterization rooms. We interviewed and collected information from officials from the Aberdeen Test Center, the U.S. Army Evaluation Center, PEO Soldier, DOD's office of the Director of Operational Test and Evaluation, and other Army components, as well as from body armor manufacturers and private body armor testing facilities. We recorded selected test data in a systematic and structured manner using a data collection instrument we developed, analyzed selected test data, and compared our observations of the way the Aberdeen Test Center conducted Preliminary Design Model testing and First Article Testing with the testing protocols that Army officials told us served as the testing standards at the Aberdeen Test Center. Specifically, these testing protocols were: (1) test procedures described in the contract solicitation announcement's purchase descriptions and (2) the Army's detailed test plans and test operations procedures that were to serve as guidance to

[6] We also issued two decisions on bid protests concerning testing under the solicitation. Armorworks Enters., LLC.,B-400394, B-400394.2, Sept. 23, 2008, 2008 CPD para. 176 (protest denied in part and dismissed in part) and Armorworks Enterprises, LLC, B-400394.3, Mar. 31, 2009, 2009 CPD para. 79 (protest dismissed).

Aberdeen Test Center testers and that were developed by the Army Test and Evaluation Command and approved by PEO Soldier, the office of the Director of Operational Test and Evaluation, Army Research Labs, and cognizant Army components. In this report, we refer to these testing standards that were to be used at Aberdeen Test Center as testing protocols. We also reviewed NIJ testing standards because Aberdeen Test Center officials told us that, although Aberdeen Test Center is not an NIJ-certified testing facility, they have made adjustments to their procedures based on those standards and consider them when evaluating Aberdeen Test Center testing practices. Complete details on our scope and methodology appear in appendix I.

We conducted this performance audit from July 2007 through October 2009 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Results in Brief

During Preliminary Design Model testing the Army took significant steps to run a controlled test and maintain consistency throughout the process but did not always follow established testing protocols and, as a result, did not achieve the intended test objective of determining which designs met performance requirements as a basis for awarding contracts. The Army's significant steps to run a controlled test included, for example, the consistent documentation of testing procedures using audio, video, and other electronic means and extensive efforts to maintain proper temperature and humidity in the test lanes. However, we identified several instances in which the Aberdeen Test Center deviated from testing protocols, including failing to test the ease of insertion of the plates into both pockets of the outer protective vest as required by the testing protocols; shooting several plates at the wrong velocity or location on the plate; and repeating failed clay calibration tests on the same block of clay—the latter having the potential to significantly affect test results. In the most consequential of the deviations from testing protocols, the Army testers incorrectly measured the amount of force absorbed by the designs tested by measuring back-face deformation at the point of aim rather than

at the deepest point of depression.[7] Army testers recognized the error after completing about a third of the test and then changed the test plan to call for measuring at the point of aim and likewise issued a modification to the contract solicitation. At least two[8] of the eight designs that passed Preliminary Design Model testing and were awarded contracts would have failed if measurements had been made to the deepest point of depression. The deviations from the testing protocols were the result of Aberdeen Test Center's incorrectly interpreting the testing protocols. Although Aberdeen Test Center officials told us that any deviations from the testing protocols required approval from PEO Soldier, the office of the Director of Operational Test and Evaluation, and other activities, in all these cases the Aberdeen Test Center procedures implemented were not reviewed or approved by officials from PEO Soldier, the Director of Operational Test and Evaluation, and other activities responsible for approving the testing protocols. Furthermore, PEO Soldier representatives were not present at Aberdeen Test Center during most of the testing, an absence that may have contributed to the fact that these deviations were not identified earlier during testing. PEO Soldier officials told us that they were not present at testing in order to ensure the independence of the testing facility, but they later acknowledged that they should have been more involved in that testing and would be more involved in future testing.[9] After concerns were raised regarding the testing conducted at Aberdeen Test Center under the May 2007 solicitation, the decision was made to not field any of the of the ESAPI and XSAPI plates awarded contracts as a

[7] In addition to stopping bullets, body armor absorbs and dissipates the force of the impact of these bullets. The amount of force absorbed is determined by measuring the depth of the depression—called back-face deformation— caused to the clay placed behind the body armor during ballistic testing: the lower the back-face deformation, the more force that is absorbed by the body armor. See figures 4 and 5 for examples of back-face deformation.

[8] After testers realized they were incorrectly measuring back-face deformation at the point of aim rather than at the deepest point, testers began to measure to both points but used the point-of-aim measure as the official measure, which according to Army officials was necessary to maintain consistency throughout testing and to not disadvantage any vendors. These two designs would have failed if the deepest point measure recorded had been used as the official measure. Because the deepest point was not measured during the first third of testing, additional designs could have improperly passed.

[9] Prior to Preliminary Design Model testing a body armor manufacturer whose design failed a prior test made public allegations that PEO Soldier had an unfair bias against its design. In an attempt to remove any appearance of bias against that manufacturer, PEO Soldier made a decision to not provide an on-site presence during Preliminary Design Model testing.

result of Preliminary Design Model testing until after First Article Testing was conducted.

During First Article Testing, while the Army addressed some of the problems identified during Preliminary Design Model testing, we observed instances in which Army testers did not follow the established testing protocols and did not maintain internal controls over the integrity and reliability of test data, raising questions as to whether the Army met its First Article Testing objective of determining whether each of the contracted armor plate designs met performance requirements. The Army resolved the problems with shot location and velocity and with the ease-of-insertion test. Also, Army technical experts from PEO Soldier who served on the Integrated Product Team were charged with testing oversight and maintained an on-site presence in the test lanes. However, Army testers continued to deviate from established testing protocols with respect to clay calibration and back-face deformation measurement as follows:

- For the clay calibration test, the Army testers followed an orally agreed-upon set of procedures that deviated from the established testing protocols. Specifically, Army testers used clay in testing that had failed the initial clay calibration test. The use of clay that has failed the calibration test could significantly impact test results. This was especially significant on a day with high failure rates when we observed clay being exposed to constant heavy, cold rain. The established testing protocols require the use of a specific type of non-hardening oil-based clay. Officials from the Army, private NIJ-certified ballistics laboratories, and the clay manufacturer told us that water exposure may contaminate the clay by changing its chemical bonding characteristics as well as by causing rapid and uneven cooling, which could affect test results. Although Army Test and Evaluation Command officials said covering the clay was not required and its exposure to water would not impact testing, these officials were unable to provide any documentation to support their position, raising concerns that exposure to rain may have impacted the testing results.

- Army testers rounded down back-face deformation measurements, which is not authorized in established testing protocols or consistent with their testing practice during Preliminary Design Model testing. Army officials said that rounding is a common industry practice and that they should have also rounded Preliminary Design Model testing results. While we did not validate this assertion, officials we spoke with from one private industry ballistics testing facility said that their practice was to always round results up, not down, which has the same

effect as not rounding. As a result of the rounding, two designs passed First Article Testing that would have failed if the measurements had not been rounded.

- The Army used a laser scanner as a new method to measure back-face deformation without adequately certifying that the scanner could function: (1) in its operational environment, (2) at the required accuracy, (3) in conjunction with its software package, or (4) without overstating deformation measurements. Army officials told us they are unable to estimate the accuracy of the laser scanner used in testing, raising concerns regarding the reliability of back-face deformation results. Aberdeen Test Center officials said they initially decided to use the laser because they did not believe it was possible to measure back-face deformations to the required level of accuracy using the digital caliper. However, officials from PEO Soldier and private NIJ-certified laboratories have told us that they believe the digital caliper method is capable of making these measurements and that the back-face deformation measurements in the testing protocols were developed using a digital caliper.[10] While it is uncertain what impact this issue had on the test results, the reliability and accuracy of the measurements may be called into question.

During First Article Testing, Army testers deviated from the established testing protocols by improperly scoring a complete penetration as a partial penetration. Army testers said they used a method to evaluate the penetration results that was discussed internally before First Article Testing but that was not described in the testing protocols or otherwise documented. As a result of this incident, one design passed First Article Testing that would have otherwise failed.[11] With respect to internal control issues, the Army did not consistently maintain adequate internal controls to ensure the integrity and reliability of test data. In one example, during ballistic testing, data were lost and testing had to be repeated because an

[10] Army and private laboratory officials told us that, on the basis of the limited data they had previously collected, they were concerned that the laser scanner may overstate back-face deformation measurements by about 2 millimeters as compared with the measurements obtained by using the digital caliper. We did not independently verify or validate the data provided by these officials. Since standards are based on measurements obtained with a digital caliper, results obtained using the laser scanner may be inconsistent/different than those obtained using the digital caliper.

[11] This design is also one of the ones that would have failed Preliminary Design Model testing had back-face deformations been measured to the deepest point as required by the testing protocols.

official accidentally hit the delete button and software controls were not in place to protect the integrity of test data. Federal internal control standards require that federal agencies maintain effective controls over information processing to help ensure completeness, accuracy, authorization, and validity of all transactions. Army officials acknowledged that before our review they were unaware of the specific internal control problems we identified. As a result of the deviations from testing protocols that we observed, three of the five designs that passed First Article Testing would not have passed under the existing testing protocols and one of the remaining two designs that passed would have failed Preliminary Design Model testing if those testing protocols had been fully followed. Thus, four of the five designs that passed First Article Testing and were certified by the Army as ready for full production would have instead failed testing at some point during the process, either during the initial Preliminary Design Model testing or the subsequent First Article testing, if the established testing protocols had been fully followed. As a result, the overall reliability and repeatability of the test results are uncertain. Although designs passed testing that would not have if the testing protocols had been followed, ballistics experts have not assessed the impact of the deviations from the testing protocols to determine if their effect is sufficient to call into question the ability of those designs to meet mission requirements. The Army has ordered 2,500 sets of plates (at two plates per set) from those vendors whose designs passed First Article Testing to be used for additional ballistics testing and 120,000 sets of plates to be put into inventory to address future requirements. However, to date, none of these designs have been fielded because, according to Army officials, there are adequate quantities of armor plates produced under prior contracts already in the inventory to meet current requirements.

To help ensure that test results are reliable, we are recommending that before any body armor plates are fielded to soldiers under the May 2007 solicitation, an assessment of the First Article Testing test data be performed by independent experts to determine whether the issues we identified had a significant effect on the test results. We are also making several recommendations intended to improve the transparency of testing by fully documenting any revised test practices so that their alignment with testing protocols is clear. Finally, we are making several recommendations to address the specific inconsistencies in test conditions we observed and to improve internal controls.

In written comments on a draft of this report, DOD generally concurred with our finding that there were deviations from the established testing protocols during Preliminary Design Model testing and First Article

Testing and with our recommendations to fully document revised test practices in the testing protocols and to improve internal controls over testing. However, DOD did not concur with our recommendation that an independent expert assessment of First Article Testing data be performed before any body armor plates are fielded to soldiers under contracts awarded under this solicitation. In the comments, DOD wrote that the deviations we identified have no significant impact on the test results and the subsequent contracting actions taken by the Army based on these test results. We disagree with DOD's assertions in this regard and continue to state that such an independent assessment is necessary to ensure that the body armor plates meet all protection requirements. We were unable to determine the full effects of deviations we observed as they relate to the quality of the armor designs and believe that such a determination should only be made based on a thorough assessment of the testing data by independent ballistics-testing experts. In light of such uncertainty and the critical need for confidence in the equipment by the soldiers, the Army would be taking unacceptable risk if it were to field these armor designs without taking additional steps to gain the needed confidence that the armor will perform as required. Consequently, we have added a matter for congressional consideration to our report suggesting that Congress consider directing DOD to either require that an independent external review of these body armor test results be conducted or require that DOD officially amend its testing protocols to reflect any revised test procedures and repeat First Article Testing to ensure that designs are properly tested. DOD's written comments are reprinted in appendix II.

Background

Army Solicitation for Body Armor

In May 2007, the Army issued a solicitation for body armor designs to replenish stocks and to protect against future threats by developing the next generation (X level) of protection. According to Army officials, the solicitation would result in contracts that the Army would use for sustainment of protective plate stocks for troops in Iraq and Afghanistan. The indefinite delivery/indefinite quantity contracts require the Army to purchase a minimum of 500 sets per design and allow for a maximum purchase of 1.2 million sets over the 5-year period.[12]

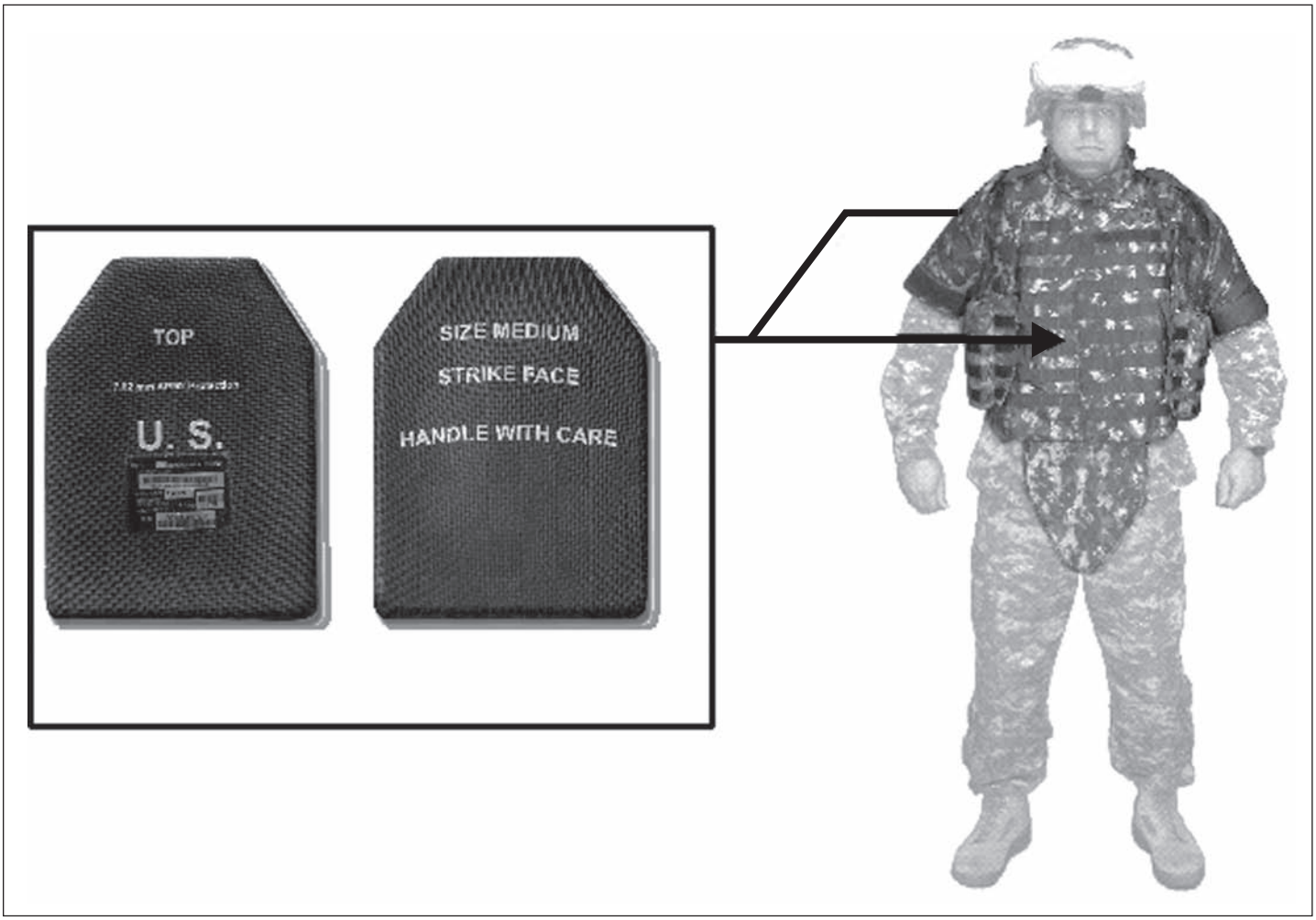
[12] A set of protective plates comprises two plates—one front and one back plate.

The Army's solicitation, which closed in February 2008, called for preliminary design models in four categories of body armor protective plates:

- Enhanced Small Arms Protective Insert (ESAPI)—plates designed to same protection specifications as those currently fielded and to fit into currently fielded Outer Tactical Vests.
- Flexible Small Arms Protective Vest-Enhanced (FSAPV-E)—flexible armor system designed to same protection specifications as armor currently fielded.
- Small Arms Protective Insert-X level (XSAPI)—next-generation plates designed to defeat higher level threat.
- Flexible Small Arms Protective Vest-X level (FSAPV-X)—flexible armor system designed to defeat higher level threat.

In figure 1, we show the ESAPI plates inside the Outer Tactical Vest.

Figure 1: ESAPI Plates as Worn inside Outer Tactical Vest



Source: Army.

Between May of 2007 and February of 2008 the Army established testing protocols, closed the solicitation, and provided separate live-fire demonstrations of the testing process to vendors who submitted items for testing and to government officials overseeing the testing. Preliminary Design Model testing was conducted at Aberdeen Test Center between February 2008 and June 2008[13] at an estimated cost of $3 million.

[13] After it was discovered that back-face deformation was being measured incorrectly, Preliminary Design Model testing was halted for 2 weeks so that Army officials could consult with senior Army leadership on how to best resolve the issue.

Additionally, over $6 million was spent on infrastructure and equipment improvements at Aberdeen Test Center to support future light armor test range requirements, including body armor testing.

First Article Testing was then conducted at Aberdeen Test Center from November 10, 2008, to December 17, 2008,[14] on the three ESAPI and five XSAPI designs that had passed Preliminary Design Model testing.[15] First Article Testing is performed in accordance with the Federal Acquisition Regulation to ensure that the contractor can furnish a product that conforms to all contract requirements for acceptance. First Article Testing determines whether the proposed product design conforms to contract requirements before or in the initial stage of production. During First Article Testing, the proposed design is evaluated to determine the probability of consistently demonstrating satisfactory performance and the ability to meet or exceed evaluation criteria specified in the purchase description. Successful First Article Testing certifies a specific design configuration and the manufacturing process used to produce the test articles. Failure of First Article Testing requires the contractor to examine the specific design configuration to determine the improvements needed to correct the performance of subsequent designs.

Testing of the body armor currently fielded by the Army was conducted by private NIJ-certified testing facilities under the supervision of PEO Soldier. According to Army officials, not a single death can be attributed to this armor's failing to provide the required level of protection for which it was designed. However, according to Army officials, one of the body armor manufacturers that had failed body armor testing in the past did not agree with the results of the testing and alleged that the testers tested that armor to higher–than–required standards. The manufacturer alleged a bias against its design and argued that its design was superior to currently fielded armor. As a result of these allegations and in response to congressional interest, after the June 2007 House Armed Services Committee hearing, the Army accelerated completion of the light armor ranges to rebuild small arms ballistic testing capabilities at Aberdeen Test

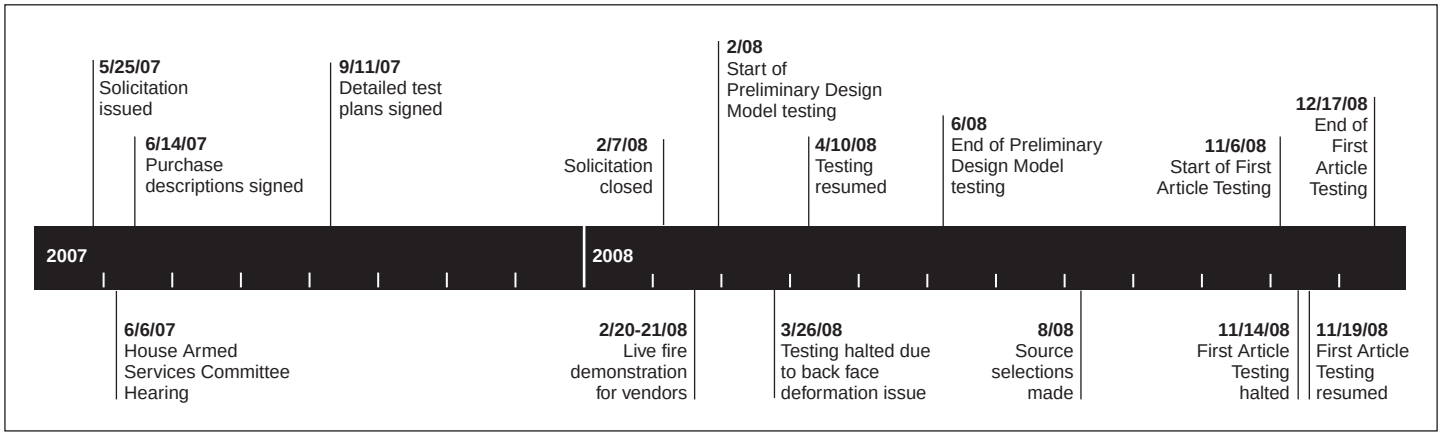
[14] Testing was halted for other high-priority tests involving 2,000 plates from Iraq that were identified as potentially cracked by nondestructive testing performed by the Army.

[15] From November 14 to November 19 First Article Testing was halted to allow for higher-priority testing to be conducted. Nearly all the ballistic testing was conducted between November 10 and December 4. The testing conducted prior to November 10 was mainly physical characterization of the plates, and the testing after December 4 was limited to the retesting of a single plate that the Army had identified as being tested incorrectly.

Center and to conduct testing under the May 2007 body armor solicitation there, without officials from PEO Soldier supervising the testing. Furthermore, the decision was made to allow Aberdeen Test Center, which is not an NIJ-certified facility, to be allowed to conduct the repeated First Article Testing. In February 2009 the Army directed that all future body armor testing be performed at Aberdeen Test Center. According to Army officials, as of this date, none of the body armor procured under the May 2007 solicitation had been fielded.

Given the significant congressional interest in the testing for this solicitation and that these were the first small arms ballistic tests conducted at Aberdeen Test Center in years, multiple defense organizations were involved in the Preliminary Design Model testing. These entities include the Aberdeen Test Center, which conducted the testing; PEO Soldier, which provided the technical subject-matter experts; and DOD's office of the Director of Operational Test and Evaluation, which combined to form the Integrated Product The Integrated Product Team was responsible for developing and approving the test plans used for the Preliminary Design Model testing and First Article Testing. Figure 2 shows a timeline of key Preliminary Design Model testing and First Article Testing events.

Figure 2: Timeline of Key Preliminary Design Model Testing and First Article Testing Events



Source: GAO observation and Army data.

Body Armor Test Procedures

The test procedures to be followed for Preliminary Design Model testing were established and identified in the purchase descriptions accompanying the solicitation announcement and in the Army's detailed

test plans (for each of the four design categories), which served as guidance to Army testers and were developed by the Army Test and Evaluation Command and approved by PEO-Soldier, DOD's office of the Director of Operational Test and Evaluation, and others. Originally, PEO Soldier required that testing be conducted at an NIJ-certified facility. Subsequently, the decision was made to conduct testing at Aberdeen Test Center, which is not NIJ-certified.[16]

The test procedures for both Preliminary Design Model testing and First Article Testing included both (1) physical characterization steps performed on each armor design to ensure they met required specifications, which included measuring weight, thickness, curvature, and size and (2) ballistic testing performed on each design. Ballistics testing for this solicitation included the following subtests: (1) ambient testing to determine whether the designs can defeat the multiple threats assigned in the respective solicitation's purchase descriptions 100 percent of the time; (2) environmental testing of the designs to determine whether they can defeat each threat 100 percent of the time after being exposed to nine different environmental conditions; and (3) testing, called V50 testing, to determine whether designs can defeat each threat at velocities significantly higher than those present or expected in Iraq or Afghanistan at least 50 percent of the time.

Ambient and environmental testing seek to determine whether designs can defeat each threat 100 percent of the time by both prohibiting the bullet from penetrating through the plate and by prohibiting the bullet from causing too deep of an indentation in the clay backing behind the plate. Preventing a penetration is important because it prevents a bullet from entering the body of the soldier. Preventing a deep indentation in the clay (called "back-face deformation") is important because the depth of the indentation indicates the amount of blunt force trauma to the soldier. Back-face deformation deeper than 43 millimeters puts the soldier at higher risk of internal injury and death.

The major steps taken in conducting a ballistic subtest include:

[16] Even though Aberdeen Test Center is not an NIJ-certified facility, Aberdeen Test Center officials said they are actively keeping abreast of NIJ standards, have made adjustments to their procedures based on those standards, and consider those standards when evaluating their own testing practices. Although there remains an active discussion in the Army testing community as to whether Aberdeen Test Center should pursue certification, Aberdeen Test Center currently has no plans to pursue NIJ certification.

1. For environmental subtests, the plate is exposed to the environmental condition tested (e.g., impact test, fluid soaks, temperature extremes, etc.).

2. The clay to be used to back the plate is formed into a mold and is placed in a conditioning chamber for at least 3 hours.

3. The test plate is placed inside of a shoot pack.

4. The clay is taken out of the conditioning chamber. It is then tested to determine if it is suitable for use[17] and, if so, is placed behind the test plate.

5. The armor and clay are then mounted to a platform and shot.

6. If the shot was fired within required specifications,[18] the plate is examined to determine if there is a complete or partial penetration, and the back-face deformation is measured.

7. The penetration result and back-face deformation are scored as a pass,[19] a limited failure,[20] or a catastrophic failure.[21] If the test is not conducted according to the testing protocols, it is scored as a no-test.

[17] Testing protocols require that clay be calibrated by dropping a 1-kilogram cylindrical weight on the clay in three locations. If all drops cause indentions between 22 and 28 millimeters, the clay is acceptable for use.

[18] Specifications include factors such as firing the shot at proper velocity, in conditions with correct humidity and temperature, and using properly conditioned clay.

[19] A "pass" is any plate that is not a limited or catastrophic failure.

[20] A "limited failure" for threats A, B, C, F, and Y is either (1) complete penetration of hard armor (the plate), but partial penetration of the soft armor (shoot pack) on any shot or (2) a back-face deformation greater than 43 millimeters but less than 48 millimeters. A limited failure on threats D and X is either (1) a complete penetration of hard armor (plate), but a partial penetration of the soft armor (shoot pack) on the first shot or (2) a complete penetration of both the hard armor (plate) and the soft armor (shoot pack) on a second shot or (3) back-face deformation on the first shot greater than 43 millimeters, but less than 48 millimeters or (4) back-face deformation greater than 43 millimeters on a second shot.

[21] A "catastrophic failure" for threats A, B, C, F, and Y is either (1) a complete penetration of the hard armor (plate) and the soft armor (shoot pack) on any shot or (2) a back-face deformation on any shot greater than or equal to 48 millimeters. A catastrophic failure for threats D and X is either (1) a complete penetration of both the hard armor (plate) and soft armor (shoot pack) on a first shot or (2) a first shot back-face deformation greater or equal to than 48 millimeters.

Army Took Significant Steps during Preliminary Design Model Testing to Run a Controlled Test and Maintain Consistency but Did Not Consistently Follow Established Testing Protocols and, as a Result, Did Not Achieve the Intended Test Objective

Army Took Significant Steps to Run a Controlled Test and Maintain Consistency

Following are significant steps the Army took to run a controlled test and maintain consistency throughout Preliminary Design Model testing:

- The Army developed testing protocols for the hard-plate (ESAPI and XSAPI) and flexible-armor (FSAPV-E and FSAPV-X) preliminary design model categories in 2007. These testing protocols were specified in newly created purchase descriptions, detailed test plans, and other documents. For each of the four preliminary design model categories, the Army developed new purchase descriptions to cover both hard-plate and flexible designs. These purchase descriptions listed the detailed requirements for each category of body armor in the solicitation issued by the Army. Based on these purchase descriptions, the Army developed detailed test plans for each of the four categories of body armor. These detailed test plans provided additional details on how to conduct testing and provided Army testers with the requirements that each design needed to pass. After these testing protocols were developed, Army testers then conducted a pilot test in which they practiced test activities in preparation for Preliminary Design Model testing, to help them better learn and understand the testing protocols.

- The Army consistently documented many testing activities by using audio, video, and other electronic means. The use of cameras and microphones to provide 24-hour video and audio surveillance of all of the major Preliminary Design Model testing activities provided additional transparency into many testing methods used and allowed for enhanced oversight by Army management, who are unable to directly observe the lanes on a regular basis but who wished to view select portions of the testing. The Army utilized an electronic database to maintain a comprehensive set of documentation for all testing activities. This electronic database included a series of data reports and pictures for each design including: physical characterization records, X-ray pictures, pre- and post-shot pictures, ballistics testing results, and details on the condition of the clay backing used for the testing of those plates.

The Army took a number of additional actions to promote a consistent and unbiased test. For example, the Army disguised vendor identity for each type of solution by identifying vendors with random numbers to create a blind test. The Army further reduced potential testing variance by shooting subtests in the same shooting lane. The Army also made a good faith effort to use consistent and controlled procedures to measure the weight, thickness, and curvature of the plates. Additionally, the Army made extensive efforts to consistently measure and maintain room temperature and humidity within desired ranges. We also observed that projectile yaw[22] was consistently monitored and maintained. We also found no deviations in the monitoring of velocities for each shot and the re-testing of plates in cases where velocities were not within the required specifications. We observed no instances of specific bias against any design, nor did we observe any instances in which a particular vendor was singled out for advantage or disadvantage.

Army Did Not Consistently Follow All Testing Protocols

We identified several instances in which the Aberdeen Test Center did not follow established testing protocols. For example, during V50 testing, testers failed to properly adjust shot velocities. V50 testing is conducted to discern the velocity at which 50 percent of the shots of a particular threat would penetrate each of the body armor designs. The testing protocols require that after every shot that is defeated by the body armor the

[22] The "yaw card" is a piece of paper placed in the intended path of the ballistic and is meant to measure the amount of yaw, or wobble, of the ballistic as it travels through the air.

velocity of the next shot be increased. Whenever a shot penetrates the armor, the velocity should be decreased for the next shot. This increasing and decreasing of the velocities is supposed to be repeated until testers determine the velocity at which 50 percent of the shots will penetrate. In cases in which the armor far exceeds the V50 requirements and is able to defeat the threat for the first six shots, the testing may be halted without discerning the V50 for the plate, and the plate is ruled as passing the requirements. During Preliminary Design Model testing, in cases in which plates defeated the first three shots, Army testers failed to increase shot velocities, but rather continued to shoot at approximately the same velocity or lower for shots four, five, and six in order to obtain six partial penetrations and conclude the test early. Army officials told us that this deviation was implemented by Aberdeen Test Center to conserve plates for other tests that needed repeating as a result of no-test events, according to Aberdeen Test Center officials—but was a practice not described in the protocols. Army officials told us that this practice had no effect on which designs passed or failed; however, this practice made it impossible to discern the true V50s for these designs and was a deviation from the testing protocols that require testers to increase velocities for shots after the armor defeats the threat.

In another example, Aberdeen Test Center testers did not consistently follow testing protocols in the ease-of-insertion test. According to the testing protocols, one barehanded person shall demonstrate insertion and removal of the ESAPI/XSAPI plates in the Outer Tactical Vest[23] pockets without tools or special aids. Rather than testing the insertion of both the front and the rear pockets as required, testers only tested the ability to insert into the front pocket. Testing officials told us that they did not test the ability to insert the plates into the rear pocket because they were unable to reach the rear pocket while wearing the Outer Tactical Vest. The cause for this deviation is that the testers misinterpreted the testing protocols, as there is no requirement in the established testing protocols to wear the Outer Tactical Vest when testing the ability to insert the plates in the rear pocket of the Outer Tactical Vest. Officials from PEO Soldier told us that, had they been present to observe this deviation during testing, they would have informed testers that the insertion test does not require that the Outer Tactical Vest be worn, which would have resulted in testers conducting the insertion test as required. According to Aberdeen Test Center officials, this violation of the testing protocols had no impact on

[23] Testing protocols use the acronym OTV for Outer Tactical Vest.

test results. While we did not independently verify this assertion, Aberdeen Test Center officials told us that the precise physical characterization measurements of the plate's width and dimensions are, alone, sufficient to ensure the plate will fit.

In addition, testers deviated from the testing protocols by placing shots at the wrong location on the plate. The testing protocols require that the second shot for one of the environmental sub-tests, called the impact test, be taken approximately 1.5 inches from the edge of the armor. However, testers mistakenly aimed closer to the edge of the armor for some of the designs tested. Army officials said that the testing protocols were unclear for this test because they did not prescribe a specific hit zone (e.g., 1.25 – 1.75 inches), but rather relied upon testers' judgment to discern the meaning of the word "approximately." One of the PEO Soldier technical advisors on the Integrated Product Team told us he was contacted by the Test Director after the plates had been shot and asked about the shot location. He told us that he informed the Test Director that the plates had been shot in the wrong location. The PEO Soldier Technical advisor told us that, had he been asked about the shot location before the testing was conducted, he could have instructed testers on the correct location at which to shoot. For 17 of the 47 total designs that we observed and measured,[24] testers marked target zones that were less than the required 1.5 inches from the plate's edge, ranging from .75 inches to 1.25 inches from the edge. Because 1.5 inches was outside of the marked aim area for these plates, we concluded that testers were not aiming for 1.5 inches. For the remaining 30 designs tested that we observed and measured, testers used a range that included 1.5 inches from the edge (for example, aiming for 1 to 1.5 inches). It is not clear what, if any, effect this deviation had on the overall test results. While no design failed Preliminary Design Model testing due to the results of this subtest, there is no way to determine if a passing design would have instead failed if the testing protocol had been correctly followed. However, all designs that passed this testing were later subject to First Article Testing, where these tests were repeated in full using the correct shot locations.[25]

[24] Nine test items could not be measured either because they were marked in a way that could not be measured or because the impact of the bullet deformed the plate too severely.

[25] One of the designs that passed Preliminary Design Model testing later failed First Article Testing because of its results during the impact test. Thus, it is possible that this design may have passed Preliminary Design Model testing due to shooting the plate at the wrong location, resulting in additional testing costs.

Of potentially greater consequence to the final test results is our observation of deviations from testing protocols regarding the clay calibration tests. According to testing protocols, the calibration of the clay backing material was supposed to be accomplished through a series of pre-test drops.[26] The depths of the pre-test drops should have been between 22 and 28 millimeters. Aberdeen Test Center officials told us that during Preliminary Design Model testing they did not follow a consistent system to determine if the clay was conditioned correctly. According to Aberdeen Test Center officials, in cases in which pre-test drops were outside the 22- to 28-millimeter range, testers would sometimes repeat one or all of the drops until the results were within range—thus resulting in the use of clay backing materials that should have been deemed unacceptable for use. These inconsistencies occurred because Army testers in each test lane made their own, sometimes incorrect, interpretation of the testing protocols. Members of the Integrated Product Team expressed concerns about these inconsistencies after they found out how calibrations were being conducted. In our conversations with Army and private body armor testing officials, consistent treatment and testing of clay was identified as critical to ensure consistent, accurate testing. According to those officials if the clay is not conditioned correctly it will impact the test results. Given that clay was used during Preliminary Design Model testing that failed the clay calibration tests, it is possible that some shots may have been taken under test conditions different than those stated in the testing protocols, potentially impacting test results. Figure 3 shows an Army tester calibrating the clay with pre-test drops.

[26] These drops comprise dropping a cylindrical metal apparatus onto the clay backing material and measuring the amount of depression caused by the drop.

Figure 3: Clay Being Calibrated with Pre-Test Drops

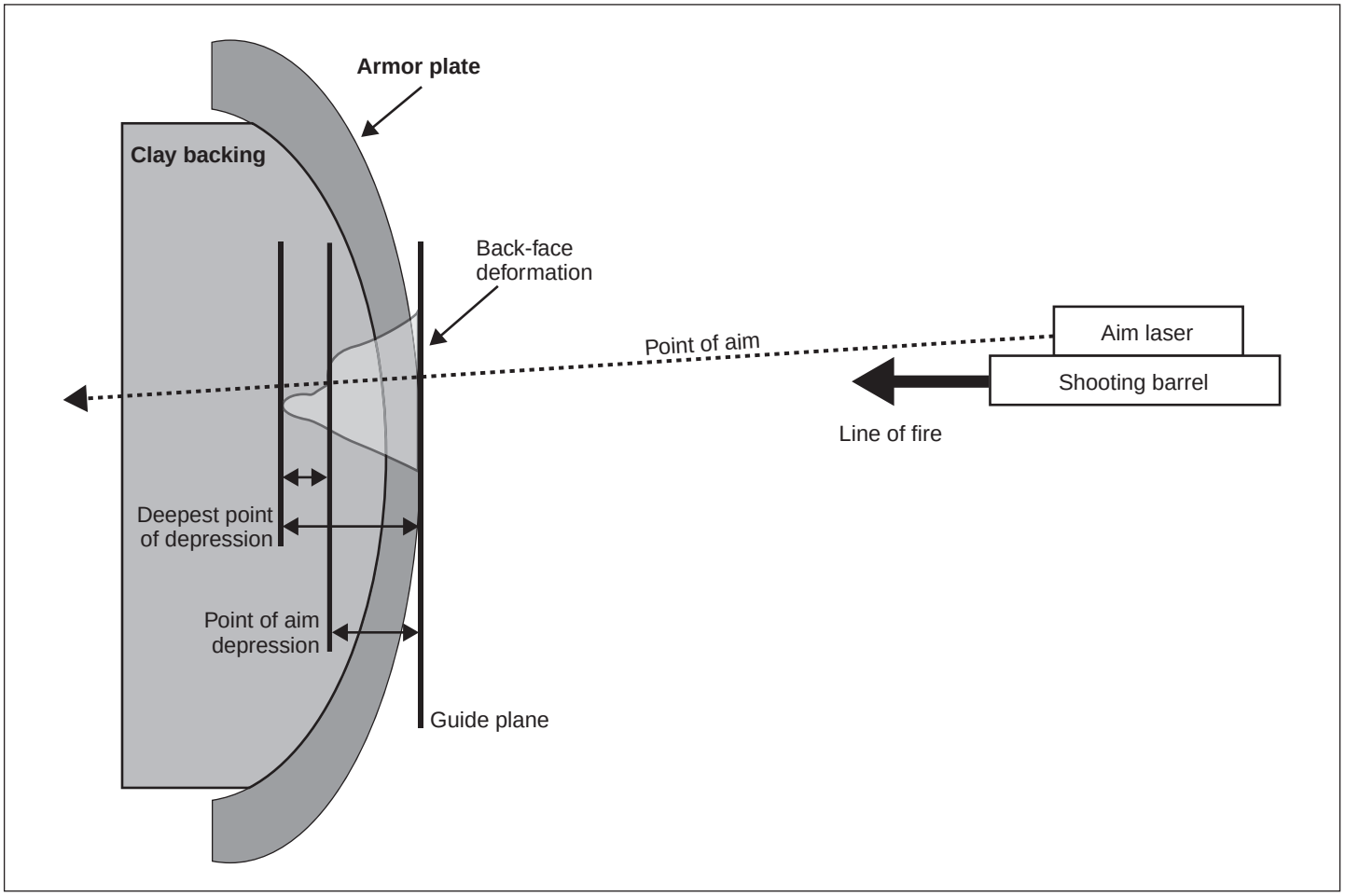
Source: Army.

The most consequential of the deviations from testing protocols we observed involved the measurement of back-face deformation, which did affect final test results. According to testing protocol, back-face

deformation is to be measured at the deepest point of the depression in the clay backing. This measure indicates the most force that the armor will allow to be exerted on an individual struck by a bullet. According to Army officials, the deeper the back-face deformation measured in the clay backing, the higher the risk of internal injury or death. During approximately the first one-third of testing, however, Army testers incorrectly measured deformation at the point of aim, rather than at the deepest point of depression. This is significant because, in many instances, measuring back-face deformation at the point of aim results in measuring at a point upon which less ballistic force is exerted, resulting in lower back-face deformation measurements and overestimating the effectiveness of the armor. The Army's subject matter experts on the Integrated Product Team were not on the test lanes during testing and thus not made aware of the error until approximately one-third of the testing had been completed.

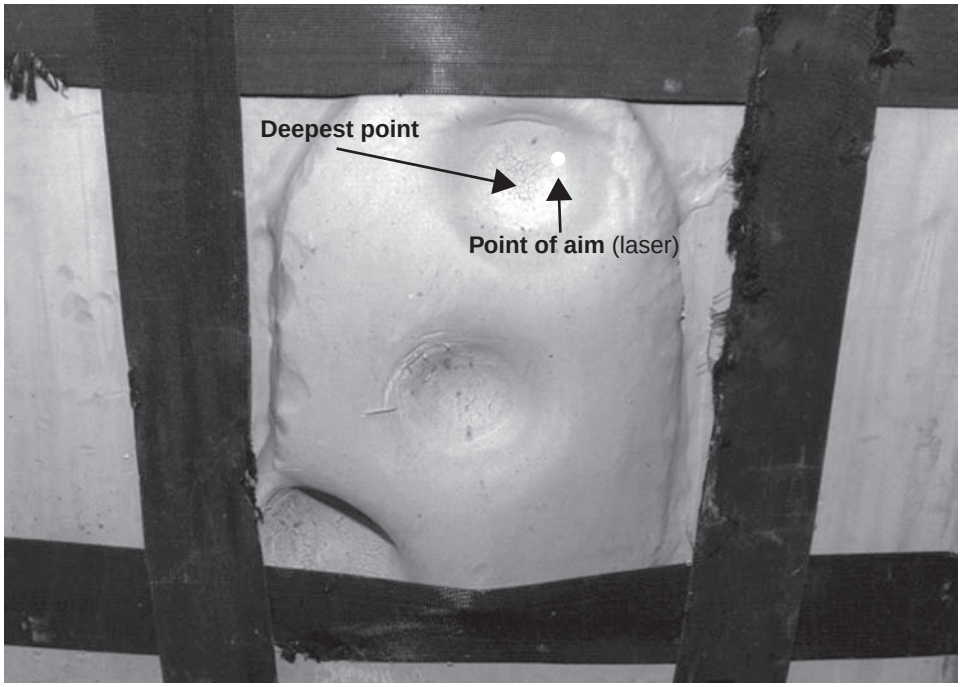
When members of the Integrated Product Team overseeing the testing were made aware of this error, the Integrated Product Team decided to begin measuring at the deepest point of depression. When senior Army leadership was made aware of this error, testing was halted for 2 weeks while Army leadership considered the situation. Army leadership developed many courses of action, including restarting the entire Preliminary Design Model testing with new armor plate submissions, but ultimately decided to continue measuring and scoring officially at the point of aim, since this would not disadvantage any vendors. The Army then changed the test plans and modified the contract solicitation to call for measuring at the point of aim. The Army also decided to collect deepest point of depression measurements for all shots from that point forward, but only as a government reference. During the second two-thirds of testing, we observed significant differences between the measurements taken at the point of aim and those taken at the deepest point, as much as a 10-millimeter difference between measurements. As a result, at least two of the eight designs that passed Preliminary Design Model testing and were awarded contracts would have failed if the deepest point of depression measurement had been used. Figures 4 and 5 illustrate the difference between the point of aim and the deepest point.

Figure 4: Graphic Representation of the Difference between the Point of Aim and the Deepest Point



Source: GAO analysis.

Figure 5: Photographic Representation of the Difference between the Point of Aim and the Deepest Point



Source: Army.

Army Decided to Repeat Testing in First Article Testing in an Attempt to Address Back-Face Deformation Measurement Problem Identified during Preliminary Design Model Testing

Before Preliminary Design Model testing began at Aberdeen Test Center, officials told us that Preliminary Design Model testing was specifically designed to meet all the requirements of First Article Testing. However, Preliminary Design Model testing failed to meet its goal of determining which designs met requirements, because of the deviations from established testing protocols described earlier in this report. Those deviations were not reviewed or approved by officials from PEO Soldier, the office of the Director of Operational Test and Evaluation, or by the Integrated Product Team charged with overseeing the test. PEO Soldier officials told us that the reason for a lack of PEO Soldier on-site presence during this testing was because of a deliberate decision made by PEO Soldier management to be as removed from the testing process as possible in order to maximize the independence of the Aberdeen Test Center. PEO Soldier officials told us that it was important to demonstrate the independence of the Aberdeen Test Center to quash allegations of bias made by a vendor whose design had failed prior testing and that this choice may have contributed to some of the deviations not being identified

by the Army earlier during testing. After the conclusion of Preliminary Design Model testing, PEO Soldier officials told us that they should have been more involved in the testing and that they would be more involved in future testing.

After the completion of Preliminary Design Model testing, the Commanding General of PEO Soldier said that, as the Milestone Decision Authority[27] for the program, he elected to repeat the testing conducted during Preliminary Design Model testing through First Article Testing before any body armor was fielded based on the solicitation. According to PEO Soldier officials, at the beginning of Preliminary Design Model testing, there was no intention or plan to conduct First Article Testing following contract awards given that the Preliminary Design Model testing was to follow the First Article Testing protocol. However, because of the fact that back-face deformation was not measured to the deepest point, PEO-Soldier and Army Test and Evaluation and Command acknowledged that there was no longer an option of forgoing First Article Testing. PEO Soldier also expressed concerns that Aberdeen Test Center test facilities have not yet demonstrated that they are able to test to the same level as NIJ-certified facilities. However, officials from Army Test and Evaluation Command and DOD's office of the Director of Operational Test and Evaluation asserted that Aberdeen Test Center was just as capable as NIJ-certified laboratories, and Army leadership eventually decided that First Article Testing would be performed at Aberdeen.

[27] The Milestone Decision Authority (MDA) is the designated individual with overall responsibility for a program. According to DOD Directive 5000.01, the MDA shall have the authority to approve entry of an acquisition program into the next phase of the acquisition process and shall be accountable for cost, schedule, and performance reporting to higher authority, including congressional reporting.

During First Article Testing the Army Addressed Some of the Problems Identified in Preliminary Design Model Testing, but Army Testers Did Not Always Follow Established Testing Protocols and Did Not Maintain Some Internal Controls

During First Article Testing, the Army Addressed Some of the Problems Identified during Preliminary Design Model Testing

PEO Soldier maintained an on-site presence in the test lanes and the Army technical experts on the Integrated Product Team charged with testing oversight resolved the following problems during First Article Testing:

- The Army adjusted its testing protocols to clarify the required shot location for the impact test, and Army testers correctly placed these shots as required by the protocols.
- After the first few days of First Article Testing, in accordance with testing protocols, Army testers began to increase the velocity after every shot defeated by the armor required during V50 testing.
- As required by the testing protocols, Army testers conducted the ease-of-insertion tests for both the front and rear pockets of the outer protective vest, ensuring that the protective plates would properly fit in both pockets.

The Army began to address the problems identified during Preliminary Design Model testing with the clay calibration tests and back-face deformation measurements. Army testers said they developed an informal set of procedures to determine when to repeat failed clay calibration tests. The procedures, which were not documented, called for repeating the

entire series of clay calibration drops if one of the calibration drops showed a failure.[28] If the clay passes either the first or second test, the clay is to be used in testing. If the clay fails both the first and the second series of drops, the clay is to then be placed back in conditioning and testers get a new block of clay. With respect to back-face deformation measurements, Army testers measured back-face deformation at the deepest point, rather than at the point of aim.[29]

Army Did Not Follow All Established Testing Protocols during First Article Testing

Although the Army began to address problems relating to the clay calibration tests and back-face deformation measurements, Army testers still did not follow all established testing protocols in these areas. As a result, the Army may not have achieved the objective of First Article Testing—to determine if the designs tested met the minimum requirements for ballistic protection. First, the orally agreed-upon procedures used by Army testers to conduct the clay calibration tests were inconsistent with the established testing protocols. Second, with respect to back-face deformation measurements, Army testers rounded back-face deformation measurements to the nearest millimeter, a practice that was neither articulated in the testing protocols nor consistent with Preliminary Design Model testing. Third, also with respect to back-face deformation measurements, Army testers introduced a new, unproven measuring device.

Although Army testers told us that they had orally agreed upon an informal set of procedures to determine when to repeat failed clay calibration tests, those procedures are inconsistent with the established testing protocols. The Army deviated from established testing protocols by using clay that had failed the calibration test as prescribed by the testing protocols. The testing protocols specify that a series of three pre-test drops of a weight on the clay must be within specified tolerances before the clay is used. However, in several instances, the Army repeated the calibration test on the same block of clay after it had initially failed until the results of a subsequent series of three drops were within the required specifications. Army officials told us that the testing protocols do not specify what procedures should be performed when the clay does not pass

[28] When asked, Aberdeen Test Center officials could not produce a memo documenting this procedure or how they knew that it was consistently applied during the test.

[29] The Aberdeen Test Center dropped the procedure that measures depression at the aim point location used during Preliminary Design Model testing.

the first series of calibration drops, so Army officials stated they developed the procedure they followed internally prior to First Article Testing and provided oral guidance on those procedures to all test operators to ensure a consistent process. Officials we spoke with from the Army, private NIJ-certified laboratories, and industry had mixed opinions regarding the practice of re-testing failed clay, with some expressing concerns that performing a second series of calibration drops on clay that had failed might introduce risk that the clay may not be at the proper consistency for testing because as the clay rests it cools unevenly, which could affect the calibration.[30] Aberdeen Test Center's Test Operating Procedure states that clay should be conditioned so that the clay passes the clay calibration test, and Army officials, body armor testers from private laboratories, and body armor manufacturers we spoke to agreed that when clay fails the calibration test, this requires re-evaluation and sometimes adjustment of the clay calibration procedures used. After several clay blocks failed the clay calibration test on November 13, 2008, Army testers recognized that the clay conditioning process used was yielding clay that was not ideal and, as a result, Army testers adjusted their clay conditioning process by lowering the temperature at which the clay was stored.

On that same day of testing, November 13, 2008, we observed heavy, cold rain falling on the clay blocks that were being transported to test lanes. These clay blocks had been conditioned that day in ovens located outside of the test ranges at temperatures above 100 degrees Fahrenheit to prepare them for testing, and then were transported outside uncovered on a cold November day through heavy rain on the way to the temperature- and humidity-controlled test lane. We observed an abnormally high level of clay blocks failing the clay calibration test and a significantly higher-than-normal level of failure rates for the plates tested on that day. The only significant variation in the test environment we observed that day was constant heavy rain throughout the day. Our analysis of test data[31] also showed that 44 percent (4 of 9) of the first shots and 89 percent (8 of 9) of the second shots taken on November 13, 2008, resulted in failure

[30] We did not conduct an independent assessment of the appropriateness of re-testing failed clay.

[31] We analyzed all V0, threat D shots. We excluded V50 shots, as well as shots from all other threats either because those tests consistently did not result in penalty points or because those threats were not tested on November 13, 2008.

penalties.[32] On all of the other days of testing only 14 percent (10 of 74) of the first shots and 42 percent (31 of 74) of the second shots resulted in failure penalties. Both of these differences are statistically significant, and we believe the differences in the results may be attributable to the different test condition on that day. The established testing protocols require the use of a specific type of non-hardening oil-based clay.[33] Body armor testers from NIJ-certified private laboratories, Army officials experienced in the testing of body armor, body armor manufacturers, and the clay manufacturer we spoke with said that the clay used for testing is a type of sculpting clay that naturally softens when heat is added and that getting water on the clay backing material could cause a chemical bonding change on the clay surface.[34] Those we spoke with further stated that the cold water could additionally cause the outside of the clay to cool significantly more rapidly than the inside causing the top layer of clay to be harder than the middle. They suggested that clay be conditioned inside the test lanes and said that clay exposed to water or extreme temperature changes should not be used. Army Test and Evaluation Command officials we spoke with said that there is no prohibition in the testing protocols on allowing rain to fall onto the clay backing material and that its exposure to water would not impact testing. However, these officials were unable to provide data to validate their assertion that exposure to water would not affect the clay used during testing or the testing results. Army test officials also said that, since the conclusion of First Article Testing, Aberdeen Test Center has procured ovens to allow clay to be stored inside test lanes, rather than requiring that the clay be transported from another room where it would be exposed to environmental conditions, such as rain.

With respect to the issue of the rounding of back-face deformation measurements, during First Article Testing Army testers did not award penalty points for shots with back-face deformations between 43.0 and 43.5 millimeters. This was because the Army decided to round back-face

[32] Testing officials disputed the inclusion of one of the plates in our analysis because it was ruled a no-test. We included this plate because we had a complete set of data for the test item and it was ruled a valid test on the lane, only to be discarded several days later because testing officials believed one of the shots was "questionable." Based on the Army's objection, we analyzed the data without this plate or two other no-test plates and found 38 percent of first shots resulted in penalties. Furthermore, our analysis revealed that the proportion of failures or penalties on November 13, 2008, still differed substantially and/or significantly from the proportion on all other days.

[33] Roma Plastilina Number 1, manufactured by Chavant, Inc.

[34] We did not independently validate the information provided by these officials.

deformation measurements to the nearest millimeter—a practice that is inconsistent with the Army's established testing protocols, which require that back-face deformation measurements in the clay backing not exceed 43 millimeters and that is inconsistent with procedures followed during Preliminary Design Model testing. Army officials said that a decision to round the measurements for First Article Testing was made to reflect testing for past Army contract solicitations and common industry practices of recording measurements to the nearest millimeter.[35] While we did not validate this assertion that rounding was a common industry practice, one private industry ballistics testing facility said that its practice was to always round results up, not down, which has the same effect as not rounding at all. Army officials further stated that they should have also rounded Preliminary Design Model results but did not realize this until March 2008—several weeks into Preliminary Design Model testing—and wanted to maintain consistency throughout Preliminary Design Model testing. The Army's decision to round measurement results had a significant outcome on testing because two designs that passed First Article Testing would have instead failed if the measurements had not been rounded.

With respect to the introduction of a new device to measure back-face deformation, the Army began to use a laser scanner to measure back-face deformation without adequately certifying that the scanner could measure against the standard established when the digital caliper was used as the measuring instrument. Although Army Test and Evaluation Command certified[36] the laser scanner as accurate for measuring back-face deformation, we observed the following certification issues:

- The laser was certified based on testing done in a controlled laboratory environment that is not similar to the actual conditions on the test lanes. For example, according to the manufacturer of the laser scanner, the scanner is operable in areas of vibration provided the area scanned and the scanning-arm are on the same plane or surface.[37] This

[35] We reviewed a few test reports for body armor testing and found instances where back-face deformation results were rounded and instances where they were not rounded.

[36] We did not evaluate the validity of the certification; however, it is worthy of note that the certification report states that the method of analysis used was somewhat unusual, that some of the results were discarded because of problems with the laser, and that changes were made to the laser during the testing process.

[37] We did not independently evaluate the manufacturer's description of the capabilities of the laser scanner.

was not the case during testing, and thus it is possible the impact of the bullets fired may have thrown the scanner out of alignment or calibration.

- The certification is to a lower level of accuracy than required by the testing protocols. The certification study says that the laser is accurate to 0.2 millimeters; however, the testing protocols require an accuracy of 0.1 millimeters or better. Furthermore, the official letter from the Army Test and Evaluation Command certifying the laser for use incorrectly stated the laser meets an accuracy requirement of 1.0 millimeter rather than 0.1 millimeters as required by the protocols. Officials confirmed that this was not a typographical error.

- The laser certification was conducted before at least three[38] major software upgrades were made to the laser, which according to Army officials may have significantly changed the accuracy of the laser. Because of the incorporation of the software upgrades, Army testers told us that they do not know the accuracy level of the laser as it was actually used in First Article Testing.

- In evaluating the use of the laser scanner, the Army did not compare the actual back-face deformation measurements taken by the laser with those taken by digital caliper, previously used during Preliminary Design Model testing and by NIJ-certified laboratories. According to vendor officials and Army subject matter experts, the limited data they had previously collected have shown that back-face deformation measurements taken by laser have generally been deeper by about 2 millimeters than those taken by digital caliper. Given those preliminary findings, there is a significant risk that measurements taken by the laser may represent a significant change in test requirements.

Although Army testing officials acknowledged that they were unable to estimate the exact accuracy of the laser scanner as it was actually used during testing, they believed that based on the results of the certification study, it was suitable for measuring back-face deformation. These test officials further stated that they initially decided to use the laser because they did not believe it was possible to measure back-face deformations to the required level of accuracy using the digital caliper. However, officials

[38] In addition, after First Article Testing was concluded Aberdeen Test Center installed additional software upgrades needed to correct errors discovered in subsequent tests. These errors, which were not identified until after First Article Testing was concluded, affected the First Article Testing results.

from PEO Soldier and private NIJ-certified laboratories have told us that they believe the digital caliper method is capable of making these measurements with the required level of accuracy[39] and have been using this technique successfully for several years. PEO Soldier officials also noted that the back-face deformation measurements in the testing protocols were developed using this digital caliper method. Army testing officials noted that the laser certification study confirmed their views that the laser method was more accurate than the digital caliper. However, because of the problems with the study that we have noted in this report, it is still unclear whether the laser is the most appropriate and accurate technique for measuring back-face deformation.

Although we did not observe problems in the Army's determination of penetration results during Preliminary Design Model testing, during First Article Testing we observed that the Army did not consistently follow its testing protocols in determining whether a shot was a partial or a complete penetration. Army testing protocols require that penalty points be awarded when any fragment of the armor material is imbedded or passes into the soft under garment used behind the plate; however, the Army did not score the penetration of small debris through a plate as a complete penetration of the plate in at least one case that we observed. In this instance, we observed small fragments from the armor three layers deep inside the Kevlar backing behind the plate. This shot should have resulted in the armor's receiving 1.5 penalty points, which would have caused the design to fail First Article Testing.[40] Army officials said that testers counted the shot as only a partial penetration of the plate because it was determined that fibers of the Kevlar backing placed behind the plate were not broken,[41] which they stated was a requirement for the shot to be counted as a complete penetration of the plate. This determination was made with the agreement of an Army subject-matter expert from PEO-Soldier present on the lane. However, the requirement for broken fibers is inconsistent with the written testing protocols. Army officials acknowledged that the requirement for broken fibers was not described in

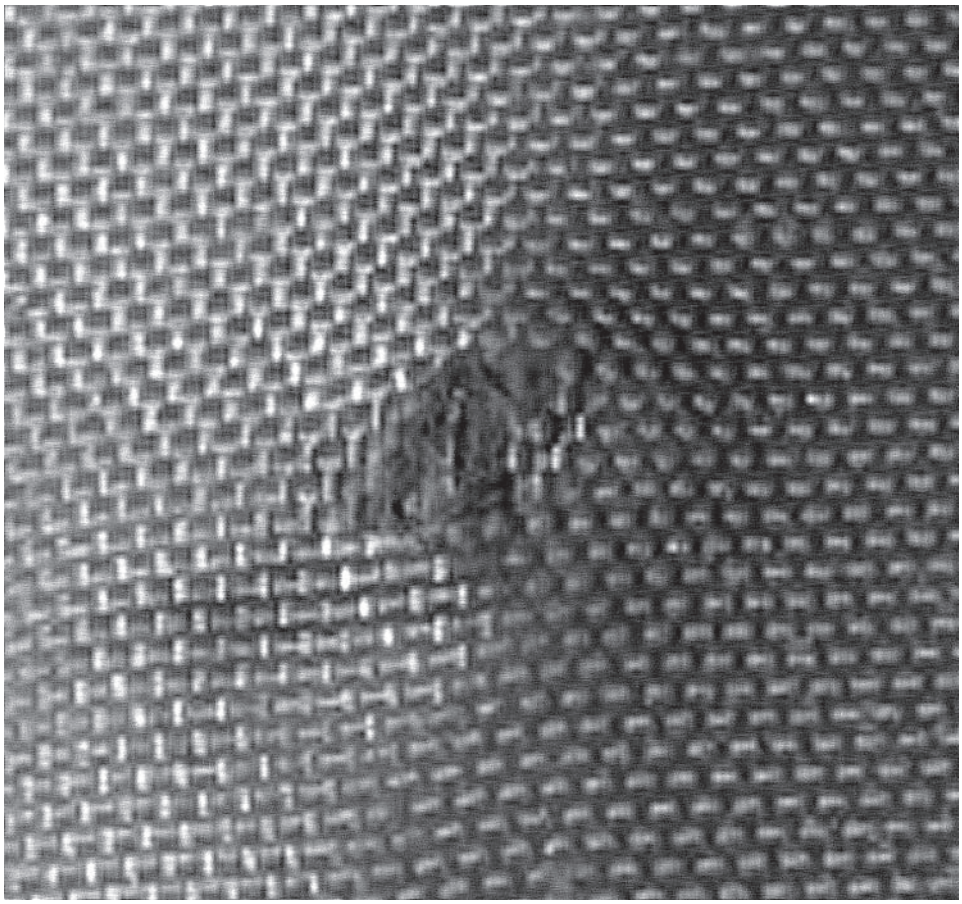
[39] We did not independently verify the level of accuracy of the digital caliper. However, the manufacturer's stated accuracy is .01 millimeters for the digital caliper specifications we obtained.

[40] This solution is also one of the ones that would have failed had back-face deformations been scored at the deepest point, rather than at the point of aim, during Preliminary Design Model Testing

[41] We observed Kevlar fibers that were frayed and tattered.

the testing protocols or otherwise documented but said that Army testers discussed this before First Article Testing began. Figure 6 shows the tear in the fibers of the rear of the plate in question.

Figure 6: Tears in Kevlar Backing Material after a Penetration of the Plate



Source: Army.

Army Did Not Maintain Internal Controls over the Integrity and Reliability of Test Data at All Times

Federal internal control standards require that federal agencies maintain effective controls over information processing to help ensure completeness, accuracy, authorization, and validity of all transactions.[42] However, the Army did not consistently maintain adequate internal controls to ensure the integrity and reliability of its test data. For example, in one case bullet velocity data were lost because the lane Test Director accidentally pressed the delete button on the keyboard, requiring a test to be repeated. Additionally, we noticed that the software being used with the laser scanner to calculate back-face deformation measurements lacked effective edit controls, which could potentially allow critical variables to be inappropriately modified during testing. We further observed a few cases in which testers attempted to memorize test data for periods of time, rather than writing that data down immediately. In at least one case, this practice resulted in the wrong data being reported and entered into the test records.

Army Did Not Formally Document Significant Procedures That Deviated from Established Testing Protocols or Assess the Impact of These Deviations

According to Army officials, decisions to implement those procedures that deviated from testing protocols were reviewed and approved by appropriate officials. However, these decisions were not formally documented, the testing protocols were not modified to reflect the changes, and vendors were not informed of the procedures. At the beginning of testing, the Director of Testing said that any change to the testing protocols has to be approved by several Army components; however, the Army was unable to produce any written documentation indicating approval of the deviations we observed by those components. With respect to internal control issues, Army officials acknowledged that before our review they were unaware of the specific internal control problems we identified. We noted during our review that in industry, as part of the NIJ certification process, an external peer review process is used to evaluate testing processes and procedures of ballistics testing facilities to ensure that effective internal controls are in place. However, we found that the Aberdeen Test Center has conducted no such reviews, a contributing factor to the Army's lack of unawareness of the control problems we noted.

As a result of the deviations from testing protocols that we observed, three of the five designs that passed First Article Testing would not have passed

[42] GAO, *Internal Control: Standards for Internal Control in the Federal Government*, GAO/AIMD-00-21.3.1 (Washington, D.C. 1999).

under the existing testing protocols. Furthermore, one of the remaining two designs that passed First Article Testing was a design that would have failed Preliminary Design Model testing if back-face deformation was measured in accordance with the established protocols for that test. Thus, four of the five designs that passed First Article Testing and were certified by the Army as ready for full production would have instead failed testing at some point during the process, either during the initial Preliminary Design Model testing or the subsequent First Article testing, if all the established testing protocols had been followed.[43] As a result, the overall reliability and repeatability of the test results are uncertain. However, because ballistics experts from the Army or elsewhere have not assessed the impact of the deviations from the testing protocols we observed during First Article Testing, it is not certain whether the effect of these deviations is sufficient to call into question the ability of the armor to meet mission requirements. Although it is certain that some armor passed testing that would not have if specific testing protocols had been followed, it is unclear if there are additional factors that would mean the armor still meets the required performance specifications. For example, the fact that the laser scanner used to measure back-face deformation may not be as accurate as what the protocol requires may offset the effects of rounding down back-face deformations. Likewise, it is possible that some of the deviations that did not on their own have a visible effect on testing results could, when taken together with other deviations, have a combined effect that is greater. In our opinion, given the significant deviations in the testing protocols, independent ballistics testing expertise would be required to determine whether or not the body armor designs procured under this solicitation provide the required level of protection. The Army has ordered 2,500 sets of plates (at two plates per set) from those vendors whose designs passed First Article Testing to be used for additional ballistics testing and 120,000 sets of plates to be put into inventory to address future requirements. However, to date, none of these designs have been fielded because, according to Army officials, there are adequate quantities of armor plates produced under prior contracts already in the inventory to meet current requirements.

[43] The one design that would have passed both the Preliminary Design Model testing and the First Article Testing actually suffered a catastrophic first-shot penalty during First Article Testing, on November 13, 2008. However, Army testers later deemed this a "questionable" shot and ruled it a no-test. The design subsequently passed its re-test.

Conclusions

Body armor plays a critical role in protecting our troops, and the testing inconsistencies we identified call into question the quality and effectiveness of testing performed at Aberdeen Test Center. Because we observed several instances in which actual test practices deviated from the established testing protocols, it is questionable whether the Army met its First Article Testing objectives of ensuring that armor designs fully met Army's requirements before the armor is purchased and used in the field. While it is possible that the testing protocol deviations had no significant net effect or may have even resulted in armor being tested to a more rigorous standard, it is also possible that some deviations may have resulted in armor being evaluated against a less stringent standard than required. We were unable to determine the full effects of these deviations as they relate to the quality of the armor designs and believe such a determination should only be made based on a thorough assessment of the testing data by independent ballistics testing experts. In light of such uncertainty and the critical need for confidence in the equipment by the soldiers, the Army would take an unacceptable risk if it were to field these designs without taking additional steps to gain the needed confidence that the armor will perform as required.

The Army is now moving forward with plans to conduct all future body armor testing at Aberdeen Test Center. Therefore, it is essential that the transparency and consistency of its program be improved by ensuring that all test practices fully align with established testing protocols and that any modifications in test procedures be fully reviewed and approved by the appropriate officials, with supporting documentation, and that the testing protocols be formally changed to reflect the revised or actual procedures. Additionally, it is imperative that all instrumentation, such as the laser scanner, used for testing be fully evaluated and certified to ensure its accuracy and applicability to body armor testing. Furthermore, it is essential that effective internal controls over data and testing processes be in place. The body armor industry has adopted the practice, through the NIJ certification program, of using external peer reviews to evaluate and improve private laboratories' test procedures and controls. This type of independent peer review could be equally beneficial to the Aberdeen Test Center. Without all of these steps, there will continue to be uncertainty with regard to whether future testing data are repeatable and reliable and can be used to accurately evaluate body armor designs. Until Aberdeen Test Center has effectively honed its testing practices to eliminate the types of inconsistencies we observed, concerns will remain regarding the rigor of testing conducted at that facility.

Recommendations for Executive Action

To determine what effect, if any, the problems we observed had on the test data and on the outcomes of First Article Testing, we recommend the Secretary of Defense direct the Secretary of the Army to provide for an independent evaluation of the First Article Testing results by ballistics and statistical experts external to DOD before any armor is fielded to soldiers under this contract solicitation and that the Army report the results of that assessment to the office of the Director of Operational Test and Evaluation and the Congress. In performing this evaluation, the independent experts should specifically evaluate the effects of the following practices observed during First Article Testing:

- the rounding of back-face deformation measurements;
- not scoring penetrations of material through the plate as a complete penetration unless broken fibers are observed in the Kevlar backing behind each plate;
- the use of the laser scanner to measure back-face deformations without a full evaluation of its accuracy as it was actually used during testing, to include the use of the software modifications and operation under actual test conditions;
- the exposure of the clay backing material to rain and other outside environmental conditions as well as the effect of high oven temperatures during storage and conditioning; and
- the use of an additional series of clay calibration drops when the first series of clay calibration drops does not pass required specifications.

To better align actual test practices with established testing protocols during future body armor testing, we recommend that the Secretary of the Defense direct the Secretary of the Army to document all key decisions made to clarify or change the testing protocols. With respect to the specific inconsistencies we identified between the test practices and testing protocols, we recommend that the Secretary of the Army, based on the results of the independent expert review of the First Article Test results, take the following actions:

- Determine whether those practices that deviated from established testing protocols during First Article Testing will be continued during future testing and change the established testing protocols to reflect those revised practices.

- Evaluate and re-certify the accuracy of the laser scanner to the correct standard with all software modifications incorporated and include in this analysis a side-by-side comparison of the laser measurements of the actual back-face deformations with those taken by digital caliper to

determine whether laser measurements can meet the standard of the testing protocols.

To improve internal controls over the integrity and reliability of test data for future testing as well as provide for consistent test conditions and comparable data between tests, we recommend that the Secretary of Defense direct the Secretary of the Army to provide for an independent peer review of Aberdeen Test Center's body armor testing protocols, facilities, and instrumentation to ensure that proper internal controls and sound management practices are in place. This peer review should be performed by testing experts external to the Army and DOD.

Matter for Congressional Consideration

DOD did not concur with our recommendation for an independent evaluation of First Article Testing results and accordingly plans to take no action to provide such an assessment. DOD asserted that the issues we identified do not alter the effects of testing. However, based on our analysis and findings there is sufficient evidence to raise questions as to whether the issues we identified had an impact on testing results. As a result, we continue to believe it is necessary to have an independent external expert review these test results and the overall effect of the testing deviations we observed on those results before any armor is fielded to military personnel. Without such an independent review, the First Article Test results remain questionable, undermining the confidence of the public and those who might rely on the armor for protection. Consequently, Congress should consider directing the Office of the Secretary of Defense to either require that an independent external review of these body armor test results be conducted or that DOD officially amend its testing protocols to reflect any revised test procedures and repeat First Article Testing to ensure that only properly tested designs are fielded.

Agency Comments and Our Evaluation

In written comments on a draft of this report, DOD takes the position that our findings had no significant impact on the test results and on the subsequent contracting actions taken by the Army. DOD also does not concur with what it perceives as our two overarching conclusions: (1) that Preliminary Design Model testing did not achieve its intended objective of determining, as a basis for contract awards, which designs met performance requirements and (2) that First Article Testing may not have met its objective of determining whether each of the contracted plate designs met performance requirements. DOD commented that it recognizes the importance of personal protection equipment such as body

armor and provided several examples of actions DOD and the Army have taken to improve body armor testing. DOD generally concurred with our findings that there were deviations from the testing protocols during Preliminary Design Model testing and First Article Testing. We agree that DOD has taken positive steps to improve its body armor testing program and to address concerns raised by Congress and others. DOD also concurred with our second recommendation to document all key decisions made to clarify or change the testing protocols. DOD did not concur with our first recommendation that an independent evaluation of First Article Testing results be performed by independent ballistics and statistical experts before any of the armor is fielded to soldiers under contracts awarded under this solicitation. Similarly, DOD did not agree with our conclusions that Preliminary Design Model testing did not meet its intended objectives and that First Article Testing may not have met its intended objectives. In supporting its position, DOD cited, for example, that rounding back-face deformation measurements during First Article Testing was an acceptable test practice because rounding is a practice that has been used historically. It was the intent of PEO Soldier to round back-face deformations for all testing associated with this solicitation, and the Integrated Product Team decided collectively to round back-face deformations during First Article Testing. However, as stated in our report and acknowledged by DOD, the rounding down of back-face deformations was not spelled out or provided for by any of the testing protocol documents. Additionally, it created an inconsistency between Preliminary Design Model testing, where back-face deformations were not rounded down and in First Article Testing, where back-face deformations were rounded down. Of greatest consequence, rounding down back-face deformations lowered the requirements that solutions had to meet to pass testing. Two solutions passed First Article Testing because back-face deformations were rounded down, meaning that the Army may be taking unacceptable risk if plates are fielded without an additional, independent assessment by experts.

DOD also did not agree with our finding that a penetration of a plate was improperly scored. DOD did agree that figure 6, which shows the tear in the Kevlar fibers of the rear of the plate in question, appears to show evidence of a perforation and that an Aberdeen Test Center ballistics subject matter expert found particles in the soft backing material behind the plate. Nevertheless, DOD did not concur with our finding because it asserted that no threads were broken on the first layer of Kevlar. However,

as we stated in the report, the protocols define a complete penetration[44] as having occurred when the projectile, fragment of the projectile, or fragment of the armor material is imbedded or passes into the soft under garment used behind the protective inserts plates, not when threads of the Kevlar are broken. The fragments found by the Aberdeen Test Center subject matter expert, as well as the three frayed, tattered, and separated Kevlar layers that we and Army testers observed, confirm our observations during testing. DOD also stated that the first layer of soft armor behind the plate under test serves as a witness plate during testing and if that first layer of soft armor is not penetrated, as determined by the breaking of threads on that first layer of soft armor, the test shot is not scored as a complete penetration in accordance with the PEO Soldier's scoring criteria. We disagree with DOD's position because the protocols do not require the use of a "witness plate" during testing to determine if a penetration occurred. If this shot would have been ruled a complete penetration rather than a partial penetration, this design would have accrued additional point deductions causing it to fail First Article Testing.

DOD did not agree that the certification of the laser scanner was inadequate and made several statements in defense of both the laser and its certification. Among these is the fact that the laser removes the human factor of subjectively trying to find the deepest point, potentially pushing the caliper into the clay, and removing the need to use correction factors, all of which we agree may be positive things. However, we maintain that the certification of the laser was not adequately performed. As indicated in the certification letter, the laser was certified to a standard that did not meet the requirement of the testing protocols. Additionally, DOD stated that software modifications added to the laser after certification did not affect measurements; however, Army testers told us on multiple occasions that the modifications were designed to change the measurements reported by the laser. DOD added that the scanner does not artificially overstate back-face deformations and relies on the verified accuracy of the scanner and the study involving the scanning of clay replicas to support its claim. Based on our observations, the scanner was certified to the wrong standard and the certification study was not performed in the actual test environment using actual shots. DOD asserts that the scanner does not overstate back-face deformations and that it does not establish a new requirement. However, DOD cannot properly validate these assertions

[44] NIJ Standard, Section 3.34 is consistent with this definition, Ballistic Resistance of Body Armor, NIJ Standard-0101.06, July 2008.

without a side-by-side comparison of the laser scanner and the digital caliper in their operational environment. Given the numerous issues regarding the laser and its certification, we maintain that its effect on First Article Testing should be examined by an external ballistics expert.

DOD also stated that it did not agree with our finding that exposure of the clay backing to heavy rain on one day may have affected test results. DOD challenged our statistical analysis and offered its own statistical analysis as evidence that it was the poor designs themselves that caused unusual test results that day. We stand by our analysis, in combination with statements made by DOD and non-DOD officials with testing expertise and by the clay manufacturer, that exposure of the clay to constant, heavy cold rain may have had an effect on test results. Further, in analyzing the Army's statistical analysis presented in DOD's comments, we did not find this information to demonstrate that the designs were the factor in unusual test results that day or that the rain exposure could not have had an effect on the results. More detailed discussions of the Army's analysis and our conclusions are provided in comments 13 and 24 of appendix II.

DOD partially disagreed that the use of an additional series of clay calibration drops when the first series of drops were outside specifications did not meet First Article Test requirements and added that all clay used in testing passed the clay calibration in effect at the time. However, we witnessed several clay calibration drops that were not within specifications. These failed clay boxes were repaired, re-dropped, and either used if they passed the subsequent drop calibration series or discarded if they failed. The protocols only allow for one series of drops per clay box, which is the methodology that Army testers should have followed. DOD stated that NIJ standards do permit the repeating of failed calibration drops. However, our review of the NIJ standards[45] reveals that there is no provision that allows repeat calibration drops. DOD states in its comments that NIJ standards are inappropriate for its test facilities, stating that these standards are insufficient for the U.S. Army given the expanded testing required to ensure body armor meets U.S. Army requirements. NIJ standards were not the subject of our review, but rather Aberdeen Test Center's application of the Army's current solicitation's protocols during testing. Further, DOD acknowledged in its comments that

[45] Army protocols require only a series of three pre-shot calibration drops. NIJ Section 4.2.5.6 requires that a series of five pre-shot and a series of five post-shot calibration drops be within specification or a new conditioned and calibration drop validated clay be used—Ballistic Resistance of Body Armor, NIJ Standard-0101.06, July 2008.

National Institute of Standards and Technology officials recommended only one series of drops for clay calibration. However, DOD stated that it will partner with the National Institute of Standards and Technology to study procedures for clay calibration, to include repeated calibration attempts, and document any appropriate procedural changes, which we agree is a good step.

Based on our analyses as described in our report and in our above responses to DOD's comments, we believe there is sufficient evidence to raise questions as to whether the issues we identified had an impact on testing results. As a result, we continue to believe that it is necessary that DOD allow an independent external expert to review these test results and the overall effect of DOD's deviations on those results before any armor is fielded to military personnel. Without such an independent review, it is our opinion that the First Article Testing results will remain questionable. Consequently, we have added a matter for congressional consideration to our report suggesting that Congress consider either directing DOD to require that an independent external review of these body armor test results be conducted or require that DOD officially amend its testing protocols to reflect any revised test procedures and repeat First Article Testing to ensure properly tested designs.

DOD partially concurred with our third recommendation to determine whether those procedures that deviated from established testing protocols during First Article Testing should be continued during future testing and to change the established testing protocols to reflect those revised procedures. DOD recognized the need to update testing protocols and added that when the office of the Director of Operational Test and Evaluation promulgates standard testing protocols across DOD, these standards will address issues that we identified. As long as DOD specifically addresses all the inconsistencies and deviations that we observed prior to any future body armor testing, this would satisfy our recommendation.

DOD stated that it partially concurs with our fourth recommendation to evaluate and recertify the accuracy of the laser scanner to the correct standard with all software modifications incorporated, based on the results of the independent expert review of the First Article Testing results. We also recommended that this process include a side-by-side comparison of the laser's measurement of back-face deformations and those taken by digital caliper. DOD concurred with the concept of an independent evaluation, but it did not concur that one is needed in this situation because according to DOD its laser certification was sufficient.

We disagree that the laser certification was performed correctly. As discussed in the body of our report and further in appendix II, recertification of the laser is critical because (1) the laser was certified to the wrong standard, (2) software modifications were added after the certification of the laser, and (3) these modifications did change the way the laser scanner measured back-face deformations. DOD did not explicitly state whether it concurred with our recommendation for a side-by-side comparison of the laser scanner and the digital caliper in their operational environment. We assert that such a study is important because without it the Army and DOD do not know the effect the laser scanner may have on the back-face deformation standard that has been used for many years and was established with the intention of being measured with a digital caliper. If the comparison reveals a significant difference between the laser scanner and the digital caliper, DOD and the Army may need to revisit the back-face deformation standard of its requirements with the input of industry experts and the medical community.

DOD generally concurred with our fifth recommendation to conduct an independent evaluation of the Aberdeen Test Center's testing protocols, facilities, and instrumentation and stated that such an evaluation would be performed by a team of subject matter experts that included both DOD and non-DOD members. We agree that in principal this approach meets the intent of our recommendation as long as the DOD[46] members of the evaluation team are independent and not made up of personnel from those organizations involved in the body armor testing such as office of the Director of Operational Test and Evaluation, the Army Test and Evaluation Command, or PEO Soldier. DOD's comments and our specific responses to them are provided in appendix II.

We are sending copies of this report to the appropriate congressional committees, the Secretary of Defense, and the Secretary of the Army. In addition, the report will be available at no charge on GAO's Web site at http://www.gao.gov. If you or your staff has any questions about this report, please contact me at (202) 512-8365 or solisw@gao.gov. Contact

[46] DOD members should not have veto power over non-DOD members.

points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made major contributions to this report are listed in appendix III.

William M. Solis
Director, Defense Capabilities and Management

List of Requesters

The Honorable Carl Levin
Chairman
The Honorable John McCain
Ranking Member
Committee on Armed Services
United States Senate

The Honorable Jim Webb
United States Senate

The Honorable Ike Skelton
Chairman
The Honorable Howard McKeon
Ranking Member
Committee on Armed Services
United States House of Representatives

The Honorable Neil Abercrombie
Chairman
The Honorable Roscoe Bartlett
Ranking Member
Subcommittee on Air and Land Forces
Committee on Armed Services
United States House of Representatives

The Honorable Joe Courtney
United States House of Representatives

Appendix I: Scope and Methodology

Our review of body armor testing focused on testing conducted by the Army in response to specific concerns raised by the House and Senate Armed Services Committees and multiple members of Congress. During our review, we were present during two rounds of testing of body armor designs that were submitted in response to a May 2007-February 2008 Army contract solicitation. The first round of testing, called Preliminary Design Model testing, was conducted from February 2008 through June 2008 with the objective of determining whether designs submitted under the contract solicitation met the required ballistic performance specifications and were eligible for contract award. The second round of testing, called First Article Testing, was conducted between November 2008 and December 2008 on the body armor designs that passed the Preliminary Design Model testing. Both tests were conducted at Aberdeen Proving Grounds in Aberdeen, Md., and were performed by Aberdeen Test Center. During the course of our review, we observed how the Army conducted its body armor testing and compared our observations with the established body armor testing protocols. We did not verify the accuracy of the Army's test data and did not provide an expert evaluation of the results of testing. To understand the practices the Army used and the established testing protocols we were comparing the practices with, we met with and/or obtained data from officials from the Department of Defense (DOD) organizations and the industry experts listed in table 1:

Table 1: Organizations Contacted for Information about Body Armor Testing

DOD acquisition organization	DOD testing organization	Industry expert
Program Executive Office Soldier	Army Test and Evaluation Command	U.S. Laboratories
	Developmental Test Command	H.P. White Laboratories
	Aberdeen Test Center	Various body armor manufacturers
	Army Research Laboratory	
	DOD's office of the Director of Operational Test and Evaluation	

Source: GAO.

To determine the degree to which the Army followed established testing protocols during the Preliminary Design Model testing of body armor designs, we were present and made observations during the entire period of testing, compared our observations with established testing protocols, and interviewed numerous DOD and other experts about body armor testing. We observed Army testers as they determined whether designs met the physical and ballistics specifications described in the contract solicitation, and as encouraged by Aberdeen Test Center officials, we

observed the ballistics testing from inside a viewing room equipped with video and audio connections to the firing lanes. We also were present and observed the physical characterization of the test items and visited the environmental conditioning chambers, the weathering chamber, and the X-ray facility. We were at Aberdeen Test Center when the designs were delivered for testing on February 7, 2008, and were on-site every day of physical characterization, which comprises the steps performed to determine whether each design meets the required weight and measurement specifications. We systematically recorded our observations of physical characterization on a structured, paper data-collection instrument that we developed after consulting with technical experts from Program Executive Office (PEO) Soldier before testing started. We were also present for every day except one of the ballistics testing, observing and collecting data on approximately 80 percent of the tests from a video viewing room that was equipped with an audio connection to each of the three firing lanes. To gather data from the day that we were not present to observe ballistic testing, we viewed that day's testing on video playback. We systematically recorded our observations of ballistics testing using a structured, electronic data-collection instrument that we developed to record relevant ballistic test data—such as the shot velocity, penetration results, and the amount of force absorbed (called "back-face deformation") by the design tested. Following testing, we supplemented the information we recorded on our data collection instrument with some of the Army's official test data and photos from its Vision Digital Library System. We developed the data collection instrument used to collect ballistics testing data by consulting with technical experts from Program Executive Office Soldier and attending a testing demonstration at Aberdeen Test Center before Preliminary Design Model testing began. After capturing the Preliminary Design Model testing data in our data collection instruments, we compared our observations of the way the Aberdeen Test Center conducted testing with the testing protocols that Army officials told us served as the testing standards at the Aberdeen Test Center. According to these officials, these testing protocols comprised the (1) test procedures described in the contract solicitation announcement's purchase descriptions and (2) Army's detailed test plans and Test Operating Procedure that serve as guidance to the Aberdeen Test Center testers and that were developed by the Army Test and Evaluation Command and approved by Program Executive Office Soldier, the office of the Director of Operational Test and Evaluation, the Army Research Labs, and cognizant Army components. We also reviewed National Institute of Justice testing standards because Aberdeen Test Center officials told us that, although Aberdeen Test Center is not a National Institute of Justice-certified testing facility, they have made adjustments to

their procedures based on those standards and consider them when evaluating Aberdeen Test Center's test practices. Regarding the edge shot locations for the impact test samples, we first measured the area of intended impact on an undisturbed portion of the test item on all 56 test samples after the samples had already been shot.[1] The next day we had Aberdeen Test Center testers measure the area of intended impact on a random sample of the impact test samples to confirm our measurements. Throughout testing we maintained a written observation log and compiled all of our ballistic test data into a master spreadsheet. Before, during, and after testing, we interviewed representatives from numerous Army entities—including the Assistant Secretary of the Army for Acquisition, Technology and Logistics; Aberdeen Test Center; Developmental Test Command; Army Research Laboratories; and Program Executive Office Soldier—and also attended Integrated Product Team meetings.

To determine the degree to which the Army followed established testing protocols during First Article Testing of the body armor designs that passed Preliminary Design Model testing, we were present and made observations during the entire period of testing, compared our observations with established testing protocols, and interviewed numerous DOD and industry experts about body armor testing. As during Preliminary Design Model testing, we observed Army testers as they determined whether designs met the physical and ballistics specifications described in the contract solicitation. However, different from our review of Preliminary Design Model testing, we had access to the firing lanes during ballistic testing. We also still had access to the video viewing room used during Preliminary Design Model testing, so we used a bifurcated approach of observing testing from both the firing lanes and the video viewing room. We were present for every day except one of First Article Testing—from the first day of ballistics testing on November 11, 2008, until the final shot was fired on December 17, 2008.[2] We noted the weights and

[1] Testers marked an area of intended impact by drawing two long lines, one marking the inner shot tolerance and the other marking the outer shot tolerance. Both our measurements and those taken by Aberdeen Test Center testers were taken by measuring the distance between the two lines and the edge on a part of the test sample significantly removed from where the shot actually impacted. We could only take these measurements on hard plate samples because the flexible samples were marked differently, in a way that we could not obtain an accurate measurement.

[2] Ballistics testing was stopped on one occasion because of a higher priority Army test that needed to be conducted at Aberdeen Test Center that involved cracked plates shipped from Iraq as part of PEO Soldier's non-destructive X-ray life cycle testing. Most of the First Article Testing concluded on December 4, but one retest was conducted on December 17.

measures of plates during physical characterization on the same data collection instrument that we used during Preliminary Design Model testing. For the ballistics tests, we revised our Preliminary Design Model testing data collection instrument so that we could capture data while in the firing lane—data that we were unable to confirm first hand during Preliminary Design Model testing. For example, we observed the pre-shot measurements of shot locations on the plates and the Aberdeen Test Center's method for recording data and tracking the chain of custody of the plates; we also recorded the depth of the clay calibration drops (the series of pre-test drops of a weight on clay that is to be placed behind the plates during the shots), the temperature of the clay, the temperature and humidity of the firing lane, the temperatures in the fluid soak conditioning trailer, and the time it took to perform tests. We continued to record all of the relevant data that we had recorded during Preliminary Design Model testing, such as the plate number, type of ballistic subtest, the charge weight of the shot, the shot velocity, the penetration results, and the back-face deformation. Regarding the new laser arm that Aberdeen Test Center acquired to measure back-face deformation during First Article Testing, we attended a demonstration of the arm's functionality performed by Aberdeen Test Center and also acquired documents related to the laser arm's certification by Army Test, Measurement, and Diagnostic Equipment activity. With a GAO senior methodologist and a senior technologist, we made observations related to Aberdeen Test Center's methods of handling and repairing clay, calibrating the laser guide used to ensure accurate shots, and measuring back-face deformation. Throughout testing we maintained a written observation log and compiled all of our ballistic test data into a master spreadsheet. Following testing, we supplemented the information we recorded on our data collection instrument with some of the Army's official test data and photos from its Vision Digital Library System to complete our records of the testing. After capturing the testing data in our data collection instruments, we compared our observations of the way Aberdeen Test Center conducted testing with the testing protocols that Army officials told us served as the testing standards at the Aberdeen Test Center. In analyzing the potential impact of independent variables on testing, such as the potential impact of the November 13th rain on the clay, we conducted statistical tests including chi-square and Fisher's Exact Test methods to accommodate small sample sizes. Before, during, and after testing, we interviewed representatives from numerous Army agencies, including Aberdeen Test Center, Developmental Test Command, Army Research Laboratories, and Program Executive Office Soldier. We also spoke with vendor representatives who were present and observing the First Article Testing, as well as with Army and industry subject matter experts.

We conducted this performance audit from July 2007 through October 2009 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Appendix II: Comments from the Department of Defense

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST
AND EVALUATION

AUG 2 9 2009

Mr. William M. Solis
Director
Defense Capabilities and Management
U.S. Government Accountability Office
441 G Street, N.W.
Washington, DC 20548

Dear Mr. Solis:

This is the Department of Defense (DoD) response to the Government Accountability Office (GAO) draft report, GAO-09-827, "WARFIGHTER SUPPORT: Independent Expert Assessment of Army Body Armor Test Results and Procedures Needed," dated July 31, 2009 (GAO Code 351282). This response includes the DoD's overall position regarding the proposed GAO report, the DoD position on each of the five GAO recommendations for executive action, and a detailed response to specific issues and assertions contained within the proposed GAO report.

The Office of the Director, Operational Test and Evaluation (DOT&E) along with the Army leadership, particularly the Office of the Assistant Secretary for Acquisition, Logistics, and Technology, Program Executive Office (PEO) Soldier, and the Army Test and Evaluation Command (ATEC), teamed in an open, collaborative, and unified manner to prepare this response. This partnership of Office of Secretary of Defense and Army leadership underscores the commitment of the DoD to develop, test, and ultimately field only the very best personal protection equipment possible. The DoD appreciates the opportunity to comment on this important report.

BACKGROUND

In preparing this response, the DOT&E, as tasked by the DoD Inspector General (IG), was the principal office of responsibility. Given that the GAO report primarily focused on testing protocols and procedures executed at the Army's Aberdeen Test Center (ATC), ATEC, parent organization of ATC, provided input to this response. As the Army's materiel developer and program manager for individual protection equipment, PEO Soldier also provided input to this response. Additionally, the U.S. Army Research, Development, and Engineering Command's Contracting Agency

participated in the review of the GAO report and provided input particularly regarding protection of source selection sensitive information in the report.

DoD recognizes the importance of personal protection equipment, the last line of defense for combat troops. In October 2003, the Acting Secretary of the Army and Army Chief of Staff directed that all measures that provide protection to Soldiers would be a focused top priority. Since that date, the Army has continually improved its personal protection products, as well as its processes associated with the procurement and testing of these items. In fact, personal protection products receive the focused attention of all Army senior leaders and over the last year its activities have been reviewed by the Secretary of the Army on a weekly basis.

In 2003, only about 10 percent of the fighting force in Iraq had ballistic inserts (hard armor plates) for their body armor. The DoD, largely led by the Army and the U.S. Marine Corps, undertook an urgent effort, working in partnership with the industrial base, to develop, procure, and field body armor plates to defeat the most significant threat in Iraq. By April 2004, about 440,000 sets of Small Arms Protective Inserts (SAPI) were fielded, equipping 100 percent of the Army and U.S. Marine Corps deployed fighting forces. Beginning about January 2005, in response to a developing threat in theater, the DoD embarked on another urgent effort to develop and field the Enhanced Small Arms Protective Insert (ESAPI). In February 2005, PEO Soldier shipped the first sets of ESAPI, and by March 2006, all Army and U.S. Marine Corps deployed warfighters were equipped with the ESAPI plates. Additionally, during this period the DoD acquired, tested, and fielded other various body armor system components in answer to warfighter needs, such as arm protection, side plates, and neck and groin protection.

See comment 1.

Undertakings of this magnitude, done urgently in time of war, are not without flaw. Inherent in this process was consideration by the DoD to incorporate into the contractual requirements, where appropriate, factors of safety above the threshold operational requirement. What was most important, however, was fielding body armor plates that defeated the threat. The DoD accomplished that objective, not once, but twice. In spite of flaws in procedures documented in recent DoD IG and Army Audit Agency reports, plates that were fielded have consistently defeated and continue to defeat the threat for which they were designed. The DoD has full confidence in the performance of the personal protection equipment its forces depend upon in combat operations.

In June 2007, the House Armed Services Committee convened a hearing to discuss allegations by a body armor manufacturer that the Army did not fairly test and evaluate its product[1]. Subsequent to that hearing and after publication of reports by the DoD IG and the GAO, the DoD has undertaken efforts to improve procedures associated

[1] On June 21, 2007, the subject company, Pinnacle Body Armor, Inc. was the subject of proposed debarment by the Department of the Air Force. On July 16, 2009, the Armed Services Board of Contract Appeals subsequently found the government's termination for cause of Pinnacle Body Armor, Inc. justified and denied its appeal.

with body armor testing. The DoD, using the Office of the DOT&E, has initiated several efforts to respond to members of the Armed Services Committees and to address issues raised by the DoD IG. Specifically, the DOT&E:

See comment 2.

- Implemented the recommendations of the DoD IG to improve the testing process via oversight and direct participation with the Services;
- Established a DoD-wide integrated project team to standardize test protocols for personal protection equipment. Included on that team are representatives from other government agencies (e.g., Federal Bureau of Investigation, Central Intelligence Agency, National Institute of Standards and Technology);
- Implemented an extensive test of body armor that will increase the statistical confidence in body armor performance and provide key input to the standardization of testing protocols;
- Implemented a process of oversight of the testing of key components of personal protection equipment;
- Established a policy to conduct personal protection equipment First Article Testing (FAT) that leads to design acceptance, at government facilities; and,
- Advised the Army and the U.S. Marines that if personal protection equipment testing is contracted to private laboratories, they should maintain government oversight during the conduct of those tests.

See comment 3.

Additionally, based on the directive of the Secretary of Defense[2], DOT&E exercised oversight of Preliminary Design Model (PDM) testing and FAT, both of which are addressed in the GAO report. DOT&E exercised this oversight by determining the scope of testing required[3], approving test plans, on-site monitoring of testing, and leading the body armor Integrated Product Team (IPT).

ACTIONS BY THE ARMY

Since 2007, ATEC has instituted procedures and policies that improve the testing of personal protection equipment. ATEC has made several investments to improve its capability and capacity for testing body armor and other personal protection equipment. The Army's investment in ATC establishes a DoD center of excellence to test personal protection equipment for all the Services. From June 2007 to the present, continued improvements have been made and are identified below.

[2] Letter to Senators Levin and McCain, dated July 12, 2007.
[3] The overall concept consists of PDM testing (Phase I), FAT, and an extended ballistic test to gather empirical data to support a new DoD standard for body armor testing (Phase II). Phase II testing is nearing completion as of the date of this communication.

3

See comment 4.

- Completed four (4) new state-of-the-art body armor test ranges with plans to construct four (4) additional test ranges within the next 18 months;
- Procured and certified a state-of-the-art laser scanner measurement device that provides accurate and repeatable measurements of Back Face Deformations (BFD);
- Developed and published Army Test Operating Procedures for testing of hard body armor that addresses Army specific requirements, which exceed those of law enforcement standards published by the National Institute of Justice (NIJ). These Army requirements address the harsh combat environmental operating conditions that Army body armor systems must endure without any degradation in performance;
- Completed installation of new clay conditioning chambers inside each test range;
- Improved velocity measurement accuracy by conducting a study of the effect of drag and creation of correction tables to more accurately capture the striking velocity of test rounds; and,
- Implemented use of electronic data collection and processing for body armor testing via the ATC Versatile Information Systems Integrated On-Line Digital Library System. Data is collected in real time and, once reviewed and authenticated, is available to authorized users over the Internet through a secure U.S. Army website. This process typically enables customers, PEO Soldier, and body armor vendors to view test results within 24 hours of test completion.

PEO Soldier has also instituted a number of efforts to improve its acquisition of personal protection equipment. These are listed below.

See comment 5.

- Ensured that all prospective body armor manufacturers may compete for new contracts;
- Transferred testing expertise and experience to support Army Acquisition Executive direction in February 2009, that all first article and lot acceptance testing would be conducted by ATEC, the Army's test agency independent of the materiel developer;
- Initiated and organized a Task Force focused on Soldier Protection that is now evolving into a new structure including a Senior Executive Service (SES) civilian managed organization with focus on quality control and procedures, decision management, process control, and compliance;
- Developed a non-destructive test capability to accurately and quickly assess ballistic plates for defects and established an in-theater, post-fielding surveillance program to examine body armor plates for cracks and other defects; and,

See comment 6.

- Developed a comprehensive and holistic personal protection evaluation process which includes pre-production, production, and post fielding activities. With this process, the Army can evaluate the effectiveness of its products at any stage of its life cycle.

The DoD recognizes the need to continually improve its procedures for the development, testing, and procurement of personal protection equipment. Many of the actions by ATEC and PEO Soldier were initiated and improved upon during the course of the GAO audit.

GENERAL COMMENTS ON THE REPORT

While the GAO Report GAO-09-827 points out some weaknesses in procedures and discrepancies in testing recently conducted by ATC, it is the DoD's position that these findings have no significant impact on the test results and the subsequent contracting actions taken by the Army based on these test results. The DoD does not concur with what it perceives as two over-arching conclusions by the GAO: 1) That Preliminary Design Model (PDM) testing did not achieve its intended objective of determining, as a basis for contract awards, which designs met performance requirements; and, 2) The FAT may not have met its objective of determining whether each of the contracted plate designs met performance requirements.

PRELIMINARY DESIGN MODEL TESTING

See comment 7.

The DoD and the Army concluded that PDM testing achieved its objective to identify those vendor designs that met the performance objectives stated in PEO Soldier's Purchase Description. The GAO cites as the Army's most consequential deviation from the test protocols described in the Purchase Description, the practice of measuring back-face deformation at the point of aim. Upon discovery of this deviation from the test protocol described in the Purchase Description, the Army stopped testing. The Army leadership, after a deliberative internal process and in consultation with DOT&E, decided to use the point of aim measurement technique as it was determined by proper authority to be an accurate and repeatable process and that it did not bias the test results against any vendor's design. The contract solicitation was modified to reflect the decision to measure at point of aim. Therefore, it is incorrect to state that the Army deviated from the test process and it is incorrect to state that "at least two" of the preliminary design models should have failed as they passed in accordance with the modified solicitation. Additional technical details supporting the rationale for this decision are found later in this letter.

See comment 8.

FIRST ARTICLE TESTING

The DoD and the Army concluded that the FAT achieved its objective of verifying that contracted vendors could produce, in a full-rate capacity, plates that passed PDM. In 2007, prior to initiation of PDM testing by ATEC, DOT&E, Army leadership, and ATEC all agreed that FAT would be conducted as part of the Army body armor testing effort that is the subject of the GAO report. Though conduct of FAT became more essential following PDM testing due to measuring BFD at point of aim as opposed to the deepest deformation, as discussed in detail in the responses to GAO's recommendations, it was never the DoD's intent to waive FAT during this effort. As a system on DOT&E oversight, it was DOT&E's responsibility to determine the required scope of testing.

The multi-phase concept that included PDM testing, FAT, and extended ballistic testing to support development of an improved test standard was briefed to congressional member and professional staff on November 14, 2007. The test plan was updated and briefed again to member and professional staff on October 27, 2008. As indicated by the GAO, PEO Soldier has in the past granted FAT waivers to vendors that submit production representative material that subsequently passes PDM testing. Though the GAO indicates that PEO Soldier may have initially contemplated FAT waivers, waivers were not permitted under the amended solicitation. Further, additional coordination by DOT&E and the Deputy Director for Land Warfare and Munitions, Under Secretary of Defense for Acquisition, Technology and Logistics on July 25, 2007, with the Military Deputy to the Army Acquisition Executive (superior of the PEO), confirmed that the Army would subject all vendors to all tests and conditions and conduct all testing at ATC. As the GAO noted, the Army refined procedures for FAT to address lessons-learned from PDM.

With regard to back-face deformation (BFD) measurement, ATEC acquired and certified a laser scanning device to accurately measure and record BFD. This device removes human interpretation from measuring a non-uniform back-face signature in clay and greatly improves measurement accuracy and repeatability. Introduction of new technology to improve accuracy beyond the existing procedures reflects ATEC's effort to improve body armor testing. This decision was approved by ATEC and senior Army leadership. It is the DoD's position that ATEC properly implemented the laser scanning instrumentation.

During FAT, ATEC and the PEO Soldier maintained open and continuous dialogue. PEO Soldier ballistic experts, the authors of many of the technical criteria in the Purchase Description, assisted in providing technical interpretations associated with scoring and ballistic phenomena.

6

With regard to rounding of the BFD measurement, ATEC, in keeping with historical procedures of PEO Soldier, instituted this process for FAT. This action was approved by the body armor IPT.

Though non-conformities did occur, FAT achieved its objective to identify those vendors that could mass produce acceptable plates. It is DoD's position that ATEC used proper scoring procedures during FAT. Therefore, it is incorrect to say that FAT did not meet its objective and it is incorrect to assert that three of five vendor designs should have failed FAT.

Comments and insights made by the GAO, as well as the DoD IG, have helped the Army refine and improve the procedures relating to body armor testing. Additionally, the DoD will continue to engage with external test and technology experts, such as the National Institute of Standards and Technology (NIST), a principal author of National Institute of Justice (NIJ) standards for law enforcement, as appropriate, during development and refinement of test procedures.

While the DoD recognizes the role independent test laboratories certified by NIJ can serve in helping meet its testing capacity needs, the DoD does not believe those laboratories should be considered external experts upon which to rely for critique of the DoD's current policies and procedures, as indicated in the GAO report. The DoD will continue to scrutinize its procedures and will pursue more open collaboration between agencies responsible for the development, acquisition, and testing of all personal protection equipment, primarily PEO Soldier, ATEC, and DOT&E. Further, the DoD will continue to accept critiques and criticisms from oversight agencies, and will continue to improve its test procedures.

RESPONSES TO GAO RECOMMENDATIONS FOR EXECUTIVE ACTION

RECOMMENDATION 1: The GAO recommends that the Secretary of Defense direct the Secretary of the Army to provide for an independent evaluation of the First Article Testing (FAT) results by ballistics and statistical experts external to DoD before any armor is fielded to Soldiers under this contract solicitation and that the Army report the results of that assessment to the Director of Operational Testing and Evaluation and the Congress. In performing this evaluation, the independent experts should specifically evaluate the effects of the following procedures observed during first article testing:

- The rounding of back-face deformation measurements.
- Not scoring penetrations of material through the plate as a complete penetration unless broken fibers are observed in the Kevlar backing behind each plate.
- The use of the laser scanner to measure back-face deformations without a full evaluation of its accuracy as it was actually used during testing, to include the use of the software modifications and operation under actual test conditions.

7

See comment 9.

- The exposure of the clay backing material to rain and other outside environmental conditions as well as the effect of high oven temperatures during storage and conditioning.
- The use of an additional series of clay calibration drops when the first series of clay calibration drops does not pass required specifications.

DOD RESPONSE: Non-Concur. The DoD does not concur with the GAO recommendation for an independent evaluation of First Article Test (FAT) results before any armor is fielded to Soldiers. The DoD's position is that the objectives of FAT were achieved; the FAT verified the vendors' ability to mass produce ballistic plates while maintaining performance standards. Anomalies identified by the GAO do not alter the results of FAT. The DoD is satisfied the FAT was properly scored despite the process discrepancies and documentation issues noted by the GAO. However, as noted below, the DoD will review test processes in general, with external assistance, which will include the areas identified below. Non-concurrence and partial concurrence pertaining to the bulleted items above are noted below, accordingly.

See comment 10.

Subject: Rounding of back-face deformation measurements.
Response: Non-Concur. The DoD does not agree that how testers recorded results by numerically rounding them should be reviewed by experts external to DoD before any armor is fielded to Soldiers under this contract solicitation. The procedure has been used historically by the National Institute of Justice (NIJ) and its certified laboratories since at least 1999. Since 1999, the Small Arms Protective Inserts (SAPI), Enhanced Small Arms Protective Inserts (ESAPI), and next generation SAPI (XSAPI) Purchase Descriptions have adopted NIJ's practice for measuring back-face deformation (BFD). Program Executive Office (PEO) Soldier established requirements of 43 mm and 48 mm for ESAPI and XSAPI BFD, as documented in the Purchase Description with the intent that testers round to the nearest whole number using ASTM E-29[4] as a guide. The DoD acknowledges that the details of the rounding practice are not adequately described in the Purchase Description. HP White, a certified NIJ lab, has historically used the same rounding rule for SAPI, ESAPI, and enhanced side ballistic inserts testing for past U.S. Army, U.S. Marine Corps, and Defense Supply Center Philadelphia contracts[5]. In November 2008, the body armor Integrated Product Team (IPT), consisting of representatives of the U.S. Army Test and Evaluation Command (ATEC), PEO Soldier, and the Office of the Director, Operational Test and Evaluation (DOT&E), agreed to use that same common rounding method and did use that technique during FAT and is using that technique for all lot acceptance tests (LATs). This practice is a policy decision that is not prohibited by any DoD

[4] American Society for Testing and Materials, ASTM E-29, Standard Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications. Approved for DoD use.
[5] The GAO noted in its report that they had reviewed past test reports and found instances of rounding and instances of not rounding.

8

or NIJ standard. As a result of the significantly smaller measurement error associated with the laser scanner, the DoD, PEO Soldier, and ATEC are reviewing the rounding methodology associated with the BFD scoring process and will make documented changes to the procedure as appropriate.

See comment 11.

Subject: Not scoring penetrations of material through the plate as a complete penetration unless broken fibers are observed in the Kevlar backing behind each plate.
Response: Non-concur. The DoD does not agree that use of the first layer of soft armor behind the plate as a witness plate should be reviewed by experts external to DoD before any armor is fielded to Soldiers under this contract solicitation. The first layer of soft armor behind the plate under test serves as a witness plate during testing. If that first layer of soft armor is not penetrated, as determined by the breaking of threads on that first layer of soft armor, the test shot is not scored as a complete penetration in accordance with the Program Executive Office (PEO) Soldier's scoring criteria. The breaking of a thread in the deeper layers of soft armor does not constitute a penetration since the stretching of material due to the force of impact could cause a thread in layers below the surface to break even though a penetration did not occur. Aberdeen Test Center (ATC) had a ballistic expert provide a preliminary score of test results. Those scores were reviewed and either agreed to or amended by three individual subject matter experts designated by the PEO Soldier as the official scorers for the test. If there is any question as to whether or not a thread is broken on the first layer, it is examined under a microscope at the discretion of the official scorers. In this case, three independent subject matter experts agreed with the ATC subject matter expert and scored the test results in question as a partial penetration. The GAO acknowledged there were no broken threads on the first layer of the soft armor. GAO also reports that its test observer discovered fragments three layers deep in the soft armor. However, the ATC subject matter expert who examined the test specimen following the test only found dust particles and some discoloration, which is not indicative of a complete penetration.

See comment 12.

Subject: The use of the laser scanner to measure back-face deformation (BFD) without a full evaluation of its accuracy as it was actually used during testing, to include the use of the software modifications and operation under actual test conditions.
Response: Non-concur. The DoD does not agree that use of a laser scanner needs to be reviewed by experts external to DoD before any armor is fielded to Soldiers under this contract solicitation. The laser scanner measurement device provides a superior tool for providing accurate, repeatable, defensible BFD measurements to the deepest point of penetration in clay. It is significant to note that it also eliminates human errors such as incorrectly selecting the location of the deepest point (a subjective decision) or piercing the clay with the sharp edge of the caliper

9

and making the depression deeper. The laser also alleviates the need to use correction factors[6] to account for the curvature of plates when making a BFD measurement that is not aligned with the path of the projectile (discussed in detail later in this letter). The technique of using correction factors results in only an approximation of the BFD measurement.

The U.S. Army Test, Measurement and Diagnostic Equipment (TMDE) Activity, Huntsville, Alabama, collaborated with Aberdeen Test Center (ATC) to conduct a study that led to the certification of the laser scanner measurement system. Certification testing was performed in both a lab environment and on the actual ranges used for testing. ATC conducted a comprehensive 3-month study of the laser scanner involving 1,920 measurements of replicas of clay forms to assess the laser's ability to accurately replicate the clay deformation.[7] Based on this study and the contributions made by TMDE, in accordance with Army Regulation 73-1, the Commanding General, Army Test and Evaluation Command (ATEC) certified the laser measurement system as approved for use for all future body armor testing. As the Army Executive Agent for testing, the implementation of the laser scanner as the most accurate tool available for measuring BFD in clay is within ATEC's mission and authority. The testing to support laser certification, the comprehensive study, and the certification by the Commanding General of ATEC all supported a Senior Army Level Integrated Product Team (IPT) decision to implement the laser as the Army-wide means of measuring BFD on clay.

Regarding use of the scanner system in the testing environment, the laser measurement system is protected in an armored enclosure to prevent damage from spall fragments. Bullets fired during testing do not impact or affect the scanner calibration. Additionally, shock effects caused by bullet impact to items under test do not affect alignment, because the alignment is independent of the testing environment. Alignment is accomplished through fixed reference points taken before each scan to ensure accuracy. The laser measurement system is calibrated twice daily per the manufacturer's instructions and will not operate until successfully calibrated after start-up.

Regarding software modifications, the software upgrades referred to in the report did not affect the measurement system in the laser scanner. The software changes made efficiencies in post-processing of the data and made enhancements to the graphical user interface of the system. These software changes had no

[6] Correction factors account for the curvature of the armor plates when making a perpendicular measurement from a reference plane. This is explained in a later section of this letter. This issue is also noted in DoD IG Report, "DoD Testing Requirements for Body Armor," dated January 29, 2009.
[7] Report #08-MS-25, "Quantum FARO Laser Scanning Body Armor Back-Face Deformation" Warfighter Directorate, Applied Science Test Division, ATC, dated September 23, 2008.

effect on the physical measurement process of BFD that was validated through the certification process.

Regarding the laser scanner "overstating" (artificially inflating) the BFD measurement, it is DoD's position that the verified accuracy of the scanner, coupled with the study involving the scanning of clay replicas, documents that the laser accurately measures the true BFD. Additionally, use of the laser scanner does not constitute the establishment of a new requirement.

The same laser scanner measurement device is widely used throughout commercial industry, to include the aeronautical industry, which has far tighter measurement tolerances than those required for body armor testing. It is also relevant to note that a similar laser arm from the same manufacturer is also in use by the National Institute of Science and Technology (NIST) for planar and non-planar measurements.

See comment 13.

Subject: The exposure of the clay backing material to rain and other outside environmental conditions as well as the effect of high oven temperatures during storage and conditioning.
Response: Non-concur. The GAO asserts that its statistical analyses indicate that exposure of the clay used to test a vendor on November 13, 2008, to "constant heavy, cold rain" caused unusual test results on that day. The DoD does not concur with that assessment. Based on the GAO assertion, the Army Test and Evaluation Command (ATEC) conducted its own statistical analyses on the test results for all designs tested on November 13, 2008.

The analyses show that the poor performance of the design in question was attributable to its marginal performance against the most formidable threat round under test, not to a brief time (seconds) of exposure of the clay to "constant heavy, cold rain." The design in question had a 70 percent failure rate during testing against the most formidable threat. Because all vendor designs were not tested on every test day, statistical analyses by test day provides far less insight than performing statistical analyses by individual design, as conducted by ATEC. Additionally, the GAO statisticians included "No Test[8]" data in their analyses. In accordance with the scoring protocols established by Program Executive Office (PEO) PEO Soldier, "No Test" data are excluded from the pool of test results and are not considered in any post-test analyses. Therefore, the statistics contained in the GAO draft report (44 percent first shot /90 percent second shot failures) are erroneous because they include invalid test data. It is also relevant to note that the Army statisticians (one from ATEC and one from the Army Research Laboratory)

[8] "No Test" events result from test anomalies such as too high of a striking velocity or impact to the test article at a location far different from the intended.

11

responsible for the review of those data are both well experienced in applying general statistical tools to body armor testing and one statistician was a major contributor to a recently released NIJ standard.[9]

Although the effects to the clay after a brief exposure to "constant heavy, cold rain" had no impact on the test results, Aberdeen Test Center (ATC) completed the planned installation of new clay conditioning chambers inside the test ranges precluding any external environmental conditioning interacting with the clay. This action improved overall test efficiency and mitigated safety risks to those handling heavy clay blocks.

Regarding high oven temperatures, the Purchase Description was modified prior to Preliminary Design Model (PDM) testing, removing the requirement for specific thermal conditioning of the clay blocks prior to calibration and subsequent testing. The purpose of thermal conditioning is to affect the clay in such a way as to promote successful calibration per the Purchase Description (three drops of a cylindrical steel mass shall each produce a deformation in the clay of 25 +/- 3 mm). PEO Soldier removed the thermal conditioning requirement, because regardless of any thermal conditioning used, the clay must pass the calibration test before it can be used for testing[10].

See comment 14.

Subject: The use of an additional series of clay calibration drops when the first series of clay calibration drops does not pass required specifications.
Response: Partially concur. The DoD concurs with the establishment of a written standard for conducting subsequent clay calibration drop tests, but non-concurs with the GAO's assertion that failed clay blocks were used during the conduct of ballistic testing at Aberdeen Test Center (ATC). **All clay used in testing passed the required clay calibration standard in effect at that time.** The National Institute of Justice (NIJ) standard[11] as verified by personnel at the National Institute of Standards and Technology (NIST), does not address specifically the issue of repeating clay calibration tests. Though NIST officials would recommend only one series of drops for clay calibration, that is not a requirement of the NIJ standard, and nothing prohibits a test activity from repeating calibration attempts on a block of clay. NIST also indicated they are not aware of any scientific studies or literature that describe how the clay properties might change as a result of performing repeated validation attempts. The DoD has agreed to partner with NIST to conduct experiments to improve the testing community's understanding of clay performance in ballistic testing. Upon completion of testing under the

[9] NIJ 0101.06, Ballistic Resistance of Body Armor, July 2008.
[10] The NIJ also removed specific thermal conditioning requirements from NIJ 0101.06, instead indicating that, "Actual conditioning temperature and recovery time between uses will be determined by the results of the validation drop test …"
[11] NIJ 0101.03, page 7, Section 5.2.9.

current Army solicitation, in coordination with NIST, the Director of Operational Test and Evaluation (DOT&E) and the Army will review the procedures for clay calibration, to include repeated calibration attempts, and document any appropriate procedural changes.

During the period of review by GAO, for reasons pertaining to the time limit established to complete ballistic testing and the concerns cited in the GAO report, ATC established and documented a revised procedure stating that only one repeat of a calibration attempt can be made. If the clay does not pass calibration upon the second attempt, it is reconditioned for later use and a new block of clay is substituted for calibration. During timed subtests, once a clay block is removed from the conditioning environment, all body armor plate testing using that clay block must be completed in 30 minutes or less or the clay block must be reconditioned and the test repeated. All clay backing material used during testing passed the calibration drop test prior to use.

RECOMMENDATION 2: The GAO recommends that the Secretary of Defense direct the Secretary of the Army to document all key decisions made to clarify or change the testing protocols.

DOD RESPONSE: Concur. The DoD recognizes the need for contemporaneous documentation and proper approvals to support any significant change to the testing protocols. The Director of Operational Test and Evaluation (DOT&E) and the Army will ensure that all key decisions made to clarify or change testing protocols be sufficiently documented. Additionally, the DoD intends to publish a series of standard personal protection equipment test protocols, beginning with soft and hard body armor. The DoD expects to publish the first of these standards by the end of this year. The standard will include detailed documentation requirements and will remediate the process discrepancies noted by the GAO.

RECOMMENDATION 3: With respect to the specific inconsistencies that the GAO identified between the test procedures and the testing protocols, the GAO recommends that the Secretary of the Army, based on the results of the independent expert review of the First Article Test results, determine whether those procedures that deviated from established testing protocols during First Article Testing will be continued during future testing and change the established testing protocols to reflect those revised procedures.

See comment 15.

DOD RESPONSE: Partially concur. It is the DoD's position that identified inconsistencies in procedures used to implement the test protocols in First Article Tests (FAT) did not alter test results. The DoD recognizes the need to update test protocols as necessitated by the adoption of new technologies and improved test procedures. The Army Test and Evaluation Command (ATEC) routinely updates Test Operating

13

Procedures[12] and participates in DoD actions to update Military Standards to ensure the latest approved test procedures are being followed. As noted above, the DoD will use the Director, Operational Test and Evaluation (DOT&E) to promulgate standard test protocols across the DoD. As reflected in the draft FY10 National Defense Authorization Act, DOT&E will ensure these standards are staffed to appropriate external agencies. Those new standards will address issues identified by the GAO.

RECOMMENDATION 4: With respect to the specific inconsistencies that the GAO identified between the test procedures and the testing protocols, the GAO recommends that the Secretary of the Army, based on the results of the independent expert review of the First Article Test results, evaluate and recertify the accuracy of the laser scanner to the correct standard with all software modifications incorporated and include in this analysis a side-by-side comparison of the laser measurements of the actual back-face deformations with those taken by digital caliper to determine whether laser measurements can meet the standard of the testing protocols.

See comment 16.

DOD RESPONSE: Partially concur. While the DoD does not concur with the GAO conclusion regarding inconsistencies and the need to recertify the laser measurement system, the DoD does concur with the concept of an independent certification of the laser measurement system and process. That process was completed prior to implementation of the laser scanner for back-face deformation (BFD) measurement. Per Army Regulation 750-43, the U.S. Army Test, Measurement and Diagnostic Equipment (TMDE) Agency, under the Assistant Secretary of the Army for Acquisition, Logistics, and Technology, is responsible for calibration of U.S. Army test instrumentation with traceability to National Institute of Standards and Technology (NIST) requirements. Following completion of this calibration, in accordance with Army Regulation 73-1, the Commanding General, Army Test and Evaluation Command (ATEC) certified the instrumentation for use during Army testing. Per documentation already provided to GAO, ATEC adhered to proper procedures and processes for certification of the laser measurement system prior to its use during testing and is in compliance with all applicable Army regulations. Software changes reported by the GAO did not affect the measurement system in the laser scanner, as indicated previously.

RECOMMENDATION 5: The GAO recommends that the Secretary of Defense direct the Secretary of the Army to provide for an independent peer review of Aberdeen Test Center's body armor test protocols, facilities, and instrumentation to ensure that proper internal controls and sound management procedures are in place. This peer review should be performed by testing experts external to the Army and DoD.

[12] Test Operating Procedures (TOPs) are formal documents published by ATEC that describe how tests are to be conducted.

See comment 17.

DOD RESPONSE: Partially concur. The DoD will conduct an independent evaluation of Aberdeen Test Center (ATC) test protocols, facilities, and instrumentation by subject matter experts for the ballistic testing of armor materiel for military applications. The DoD is in discussion with the National Institute of Standards and Technology (NIST) to form a team of subject matter experts to review the DoD's testing procedures. This review will be broad and will include measurement processes, clay conditioning, and other areas as appropriate. The DoD will include experts from within the DoD as part of this team.

DETAILED COMMENTS KEYED TO THE GAO REPORT

Part 1: GAO Observations and Conclusions on Preliminary Design Model Testing.

See comment 18.

Assertion 1 – Army's Aberdeen Test Center (ATC) had never before performed testing on body armor plates. (Page 2 of 48, paragraph 1; Page 13 of 48, paragraph 1; and, Page 36 of 48, paragraph 2.)
Response: ATC did the initial testing on the Interceptor Body Armor system in the 1990s and has been extensively involved in body armor long before and since that time. While ATC did not perform any additional testing on the Interceptor Body Armor system for Program Manager Soldier Equipment since that initial testing, ATC has consistently performed required body armor testing for the U.S. Special Operations Command, as well as research and development testing on body armor for organizations such as the Natick Soldier Safety Center and other service Program Managers. ATC tested over a dozen different hard armor plate designs between 1997 and 2007, to include both Small Arms Protective Inserts (SAPI) and Enhanced Small Arms Protective Inserts (ESAPI) plates.

See comment 19.

Assertion 2 – Based on Preliminary Design Model (PDM) Test results, the Army awarded contracts totaling over $8 Billion for production of ESAPI and XSAPI. (Page 2 of 48; paragraph 2.)
Response: The Army awarded three Indefinite Delivery, Indefinite Quantity (IDIQ) contracts for ESAPI and XSAPI ballistic plates. To date, those contracts have obligated $119,703,145.49 for XSAPI and $1,756,044.80 for ESAPI. The contractually guaranteed quantities on each of the three contracts have been satisfied.

See comment 20.

Assertion 3 – Aberdeen Test Center shot several plates at the wrong location on the plate. (Page 5 of 48, paragraph 1; and, Page 19 of 48, paragraph 1.)
Response: ATC followed the Purchase Description protocol for shot location during PDM testing and FAT. Specifically, the second test shot location for the impact subtest during PDM testing, as stated in the Purchase Description, was approximately 1.5 inches from an edge of the plate. There were no limits or range specified for this second test shot location. Given the potential variances between the actual aim point and impact point during testing, the tester interpreted 1.0 inch to be an acceptable aim point location

for this subtest. In this case, shooting closer to the edge would have increased the risk of a failure for this subtest, but no vendors failed testing as a result of the tester's interpretation of the second test shot location. Therefore, there was no impact on the outcome of the test.

See comment 21.

Assertion 4 – Aberdeen Test Center shot several plates at the wrong velocity. (Page 5 of 48, paragraph 1; and, Page 17 of 48, paragraph 3.)
Response: During V_{50} testing, it is worthwhile to note that one of the test threats, threat "c," is not robust enough to achieve a complete penetration no matter how much the velocity is increased; therefore, following the test procedure to achieve a complete penetration is an impossible task for threat "c."[13] However, in accordance with the PDM test protocol, the threat "c" V_{50} testing was completed to a degree that provided the required government reference data for baseline comparison to data generated during testing of previous generations of body armor. The V_{50} subtests for more robust threats, during which complete penetrations were achievable, were executed to the standard protocols.

See comment 22.

Assertion 5 – Army tester's incorrectly measured the amount of force absorbed by the designs tested by measuring back-face deformation at the point of aim rather than at the deepest point of depression. (Page 5 of 48, paragraph 1; and, Page 21-23 of 48.)
Response: Though ATC deviated from the test protocol in the Purchase Description regarding measuring BFD at the lowest deformation, Army leadership after a deliberative internal process and in consultation with DOT&E, decided to use the BFD point-of-aim measurement as it was determined to be an accurate and repeatable process that did not bias any vendor's design. The following information highlights the rationale for this decision. For clarity and reference, Figures 1 and 2 depict the geometry associated with this issue. Further discussion follows the figures.

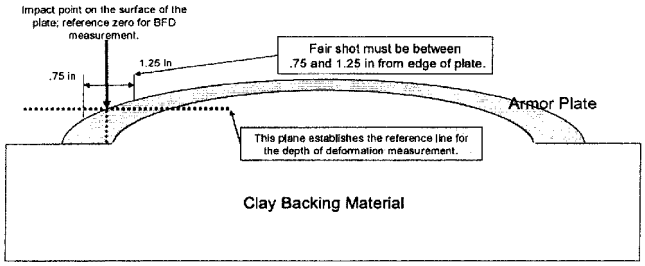


Figure 1. Test setup before test shot.

[13] Testing with threat "c" is nonetheless required to ensure that the ballistic plate defeats all threats it is designed to defeat.

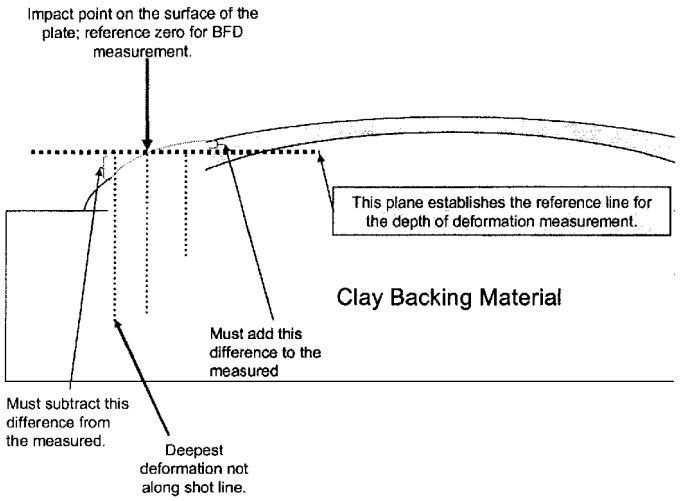


Figure 2. Post shot measurement.

The Purchase Description, Paragraph 4.9.9.3 Back-Face Deformation Measurement, states:

> "Back-face deformations in the clay will not exceed 1.70-inches (43mm) max (Paragraph 3.9.3) when measured from the original undisturbed surface of the backing material to the lowest point of the depression. All Back-Face Deformation measurements will be conducted at 0 degree obliquity only. Indentation measurements will utilize measurement devices (+/-0.1 mm accuracy) incorporating a fixed reference "guide" (See Figure 2)[14] that can rest solidly upon two edges of the fixture, establishing the reference plane across the diameter of the indentation. The distance between the reference "guide" and original undisturbed surface will be measured at the point of intended impact prior to impact. The distance between reference "guide" and the lowest point of depression will be measured after impact. Back-face deformation will be the difference between the two."

Referencing the first sentence of the above quote, the only known reference point prior to a test shot is the aim point ("original undisturbed surface of the backing material"). Prior to the shot, the test technician has no way of knowing the surface location perpendicular to the lowest point of depression. The fixed reference guide identified in the quote, shown in Figure 3, can only assist in establishing the aim point as

[14] Figure 3 in this letter is the "Figure 2" cited in the quote from the Purchase Description.

the reference point from which post-test measurements can be made. Therefore, the ATC interpretation was to use that point to measure BFD following the test.

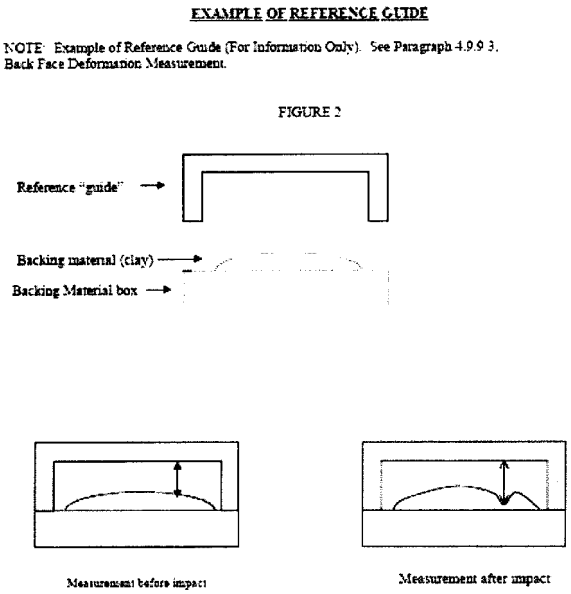


Figure 3. BFD Measurement Reference Guide from Purchase Description.

There are two important considerations pertaining to the guidance in the Purchase Description: First, the lowest point of depression must be subjectively selected by a test technician, and second, measuring perpendicular from the point selected as the lowest point of depression to the reference guide provides the distance between that point and the original undisturbed surface at the aim point, not the original undisturbed surface perpendicular to the lowest point of depression, which is what is required to obtain an accurate measurement. Using the technique described, an accurate measurement can only be made if the measurement is made on a flat surface with a uniform original undisturbed reference surface. However, since the plates are curved and the radius varies between vendor plates and even within vendors for same plate sizes, as shown in Figure 4, the only point at which an accurate measurement can be made with the described measuring device is the aim point.

18

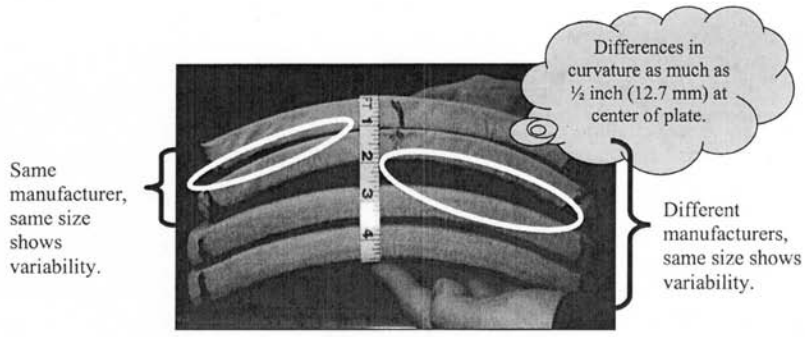


Figure 4. Picture illustrates variance in curvature between same size ballistic plates.

Although not contained in the Purchase Description, PEO Soldier had an internally documented process to account for plate curvature when the deepest point of deformation was laterally offset from the point of aim. However, due to plate curvature variance, a correction factor only approximates the deepest deformation measurement. That is one reason ATC elected to measure BFD at point of aim at the beginning of testing, but also became the rationale for a better procedure using the laser scanner as explained elsewhere in this letter.

During the early stages of PDM testing, ATC analyzed the effect of measuring BFD at the aim point in lieu of the lowest point of depression with a caliper. The results of the analysis for those 222 test data points are summarized below.

- Mean delta between aim point and deepest point BFD was 0.45 mm.
- For the 222 test shots, the minimum delta was 0 mm and the maximum was 5.95 mm.
- 170 of 222 data points showed no difference, i.e., aim point same as lowest point of depression within the BFD.
- Approximately 24 percent of measurements (52) showed a difference. i.e., aim point different than lowest point of depression.

When Army leadership and DOT&E learned of this issue of measuring BFD at aim point versus deepest deformation, ATEC halted testing. Following analysis of available data, the Army Acquisition Executive, the PEO Soldier, the Commanding

19

General of ATEC, and a senior representative from DOT&E agreed that for the remainder of PDM testing, ATEC would score BFD at the aim point, while recording for government reference the BFD at the lowest point using the aforementioned curvature correction standard. This decision is also referenced in the below executive summary from the acting Army Acquisition Executive's office following the Senior Army Leadership Executive IPT meeting:

> "Body Armor (ESAPI and XSAPI) Testing (U). (SAAL-SMS) On 26 March, the Army Test and Evaluation Command Commander suspended testing pending discussion with Army leaders from the acquisition and test communities on test procedures. The Army Acquisition Executive (AAE) convened a meeting on 07 April 2008 to explore the solicitation language and test procedures concluding that Aberdeen Test Center methods were applied consistently to all vendors and remain defendable. The AAE directed Program Executive Officer Soldier, in collaboration with the test community, to amend the solicitation to clarify the description of this procedure. The source selection and resultant contract awards will be made in accordance with the terms and conditions of the solicitation. As previously briefed to the staff of the HASC and SASC, a phase II, First Article Test, will commence later this year. This test, conducted by ATEC, will use government provided test articles purchased from the contract awardees and will encompass both ballistic and operational testing. The Research, Development, and Engineering Center Contracting Agency subsequently modified the solicitation (Amendment 14, dated April 17, 2008) to evident the Army's decision to measure BFD at point of aim for PDM testing."

At PDM completion, ATC analyzed the 3,404 data points (1,702 shots; two measurements per shot) of the two measurement methods. The results of that analysis are below.

- For 1,091 of the 1,702 test shots (64 percent), the aim point was the same as the deepest point.
- For 611 of the 1,702 test shots (36 percent), the deepest point did not occur at the aim point.
- The mean difference (deepest minus aim point) of all 1,702 test shots was 0.60 mm.
- The mean difference of only the 611 test shots with a non-zero difference between aim point and deepest point was 1.67 mm.
- For the 1,702 test shots, the minimum delta was -0.28 mm and the maximum was 10.66 mm.

20

These data show that while there is a difference in depth between the BFD measured at aim point and at lowest point, the difference is small. Nonetheless, the Army's adoption of the laser scanner measurement technique resolves this issue completely.

See comment 23.

Assertion 6 – Deviations from the test protocols (e.g., measuring back-face deformation at the aim point) were not reviewed or approved by officials from PEO Soldier, Director of Operational Test and Evaluation, and other activities responsible for approving testing protocols. (Page 5-6 of 48.)
Response: The DoD acknowledges these shortcomings.

- The DoD acknowledges that measuring BFD at the point of aim during the early stages of PDM testing was not known by all members of the IPT or senior Army or DOT&E leaders. However, once that issue became evident, all members of the IPT and their leadership acted decisively to arbitrate, resolve, and document the resolution.
- The issue of rounding BFD was discussed at an IPT meeting and was agreed to by all members present.
- ATC developed internal procedures for clay calibration that at first were not documented adequately. However, once the issue became evident, ATC proceeded to adequately document and discuss its procedures with the IPT.

These issues have been subsequently addressed. It is the DoD position that none of these issues prevented the Army from achieving its PDM objectives.

Part 2: GAO Observations and Conclusions on First Article Testing.

See comment 24.

Assertion 7 – Calibration test procedures deviated from established test protocols. GAO observed clay being exposed to "constant heavy, cold rain." (Pg 6 of 48 paragraph 3; Page 20 of 48, paragraph 1; Page 27 of 48, paragraph 2; and, Page 38 of 48, paragraph 1.)
Response. This discussion appends and adds detailed information to the discussion on this same subject contained in an earlier section. Clay is largely impervious to water penetration and the first step in preparing the clay for the calibration drop test is to scrape off the top layer of clay. This action cleared the clay block of any residual water. Regardless of the exposure to environmental conditions, the test standard is clear: If the clay backing material passes the calibration drop test it is acceptable for use for ballistic testing.

ATC analysis also shows that the poor performance of the vendor's design in question was attributed to its marginal performance against the most formidable threat round under test, not to a brief time (seconds) of exposure to a "constant heavy, cold rain and low temperatures" or the effects of rain on the clay surface. Statistical analyses were performed on the test results for all designs including those tested on November 13, 2008. The result of that analysis, as shown in Figure 5, indicate Design K was the weakest

design on days with no rain as well as days with rain. Design K had a 70 percent failure rate during all testing against the most formidable threat. Even when excluding the rain day data, which was when that design was subjected to the most stressing tests, the design still had a 57 percent failure rate.

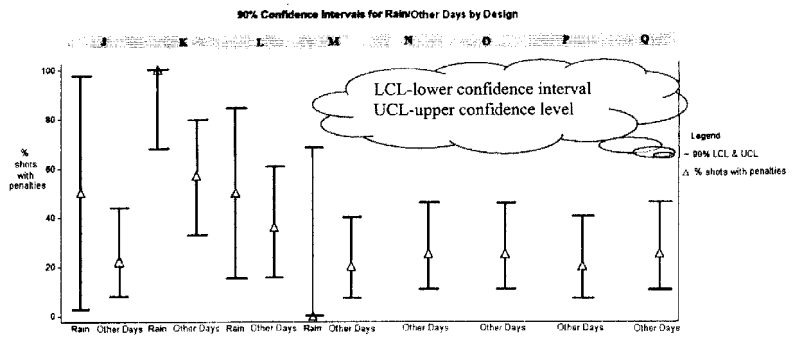


Figure 5. Results of ATC analysis of test results for all vendor designs against the most formidable threat.

In the report, GAO statisticians stated that "No Test" data was included in the statistical calculations reported on page 13 of the draft GAO report. In accordance with the scoring protocols established between the PM and ATEC and included in the Purchase Description, "No Test" data is excluded from the database and does not constitute a "failure" since it is an unfair test shot impact that requires a retest. Therefore, the statistics contained in the GAO draft report (44 percent first shot / 90 percent second shot failures) are erroneous because they include invalid test data.

See comment 25.

Assertion 8 - During First Article Testing Army testers improperly scored a complete penetration as a partial penetration. (Page 8 of 48, paragraph 1; and, Page 32 of 48, paragraph 2.)

Response: This information appends that provided in the earlier discussion of this issue. While Figure 6 of the GAO report appears to show evidence of a perforation on the rear of the test plate in question, by the definition in the Purchase Description, a complete perforation is scored based upon the damage to the soft armor shoot pack directly behind the ballistic test plate. The front (face immediately behind the plate under test) of the shoot pack in question is shown in Figure 6 of this response, located below. Though deformed, the damage does not constitute a complete perforation of the plate according to the definition in the Purchase Description.

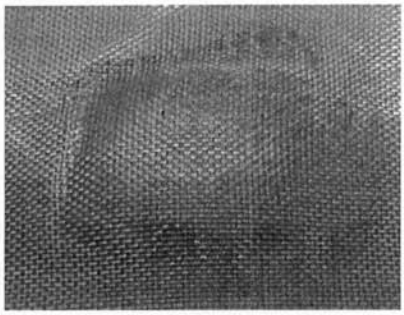


Figure 6. Front face of shoot pack for the test in question.

See comment 26.

Assertion 9 – The Army did not maintain internal controls over the integrity and reliability of test data at all times. (Page 8 of 48, paragraph 1; and, Page 33 of 48, paragraph 1.)
Response. ATEC did maintain adequate internal controls to ensure the integrity and reliability of test data. GAO reports the incident of a Test Director accidentally pressing the delete key on a computer keyboard, thereby resulting in loss of data. That did occur, but only once in over 4,100 test events. Additionally, as it was obvious the loss of data occurred, the test shot was immediately repeated. This had no effect on the outcome of the test. ATC employs a multiple, redundant system of human checks to ensure data is accurately recorded, whether recorded initially by human or computer. All test shot data is immediately uploaded to an archival system. At the end of each day, a team of analysts scrutinize those data to ensure accuracy. Only then do those data become authenticated. This is a rigorous process that helps to ensure accuracy and integrity in data acquisition, recording, and archiving. Regarding the laser scanner, only two persons are authorized and able to modify the laser scanner software. Range personnel cannot alter the scanner software or settings.

See comment 27.

Assertion 10 – The Army did not formally document significant procedure changes that deviated from established testing protocols or assess the impact of these deviations. (Page 34 of 48, paragraph 2.)
Response. Acknowledged. As indicated previously, these shortcomings have been identified, partly by the GAO, and have been remedied.

Part 3: GAO Observations and Other Issues.

Assertion 11 – GAO statement that the requirement to test at an NIJ certified test lab was withdrawn because Aberdeen Test Center is not NIJ certified. (Page 12 of 48; paragraph 2; and, Page 14 of 48; paragraph 1.)

23

See comment 28.

Response: The DoD does not believe that NIJ certification is appropriate for its test facilities. There are significant differences between NIJ and U.S. Army body armor test requirements. The NIJ certification process is intended to ensure domestic police forces and Justice Department personnel are provided body armor that meets appropriate standards for their job duties. However, these standards are insufficient for the U.S. Army, given the expanded testing required to ensure body armor meets U.S. Army requirements. Figure 7 depicts the differences between NIJ and Army body armor test requirements.

National Institute of Justice Standard 0101.04 versus Army ESAPI First Article Test

Body Armor Test	NIJ Standard	Army Standard
Resistance to Penetration Test (V0)	√	√
Ballistic Limit Test (V50)	√	√
3-Minute Water Spray Resistance to Penetration Test	√	X*
Flammability Resistance to Penetration Test (250 deg F)	X	√
Oil Immersion Resistance to Penetration Test	X	√
Diesel Immersion Resistance to Penetration Test	X	√
2-Hour Salt Water Immersion Resistance to Penetration	X	√
Impact Drop Resistance to Penetration Test	X	√
Low Temperature Resistance to Penetration Test (-60 deg F)	X	√
High Temperature Resistance to Penetration Test (160 deg F)	X	√
Temperature Shock Resistance to Penetration Test (-25 to 120 deg F)	X	√
Weatherometer/Accelerated Aging Resistance to Penetration Test	X	√
Altitude Resistance to Penetration Test	X	√

*2-Hour Salt Water Immersion Test conducted by Army is more stringent than the NIJ 3-minute Water Spray Test. **Note: X = not required by NIJ, though NIJ-certified laboratories could execute or contract out these tests.**

Figure 7. Comparison of the NIJ Standard and U.S. Army Performance Tests.

See comment 29.

Assertion 12 – After the June 2007 House Armed Services Committee hearing, the Army decided to rebuild small arms ballistic testing capabilities at Aberdeen Test Center. (Page 12 of 48; paragraph 2.)

Response: This assertion is incorrect. The contract to construct additional test ranges at the Aberdeen Test Center Light Armor Range was already awarded (September 2006) and construction was already underway at the time of June 2007 House Armed Services Committee hearing. This upgrade was not in response to any particular event, but was undertaken to meet projected future Army ballistic test requirements.

David W. Duma
Acting Director

25

GAO Comments

1. The Department of Defense (DOD) stated that undertakings of this magnitude are not without flaws and that what was most important was fielding body armor plates that defeated the threat. While DOD may have identified some flaws that may not be serious enough to call the testing results into question, several of the deviations to the testing protocols that we observed do call the testing results into question for the reasons stated in our report. An independent expert has not evaluated the impact of these deviations on the test results and, until such a study is conducted, DOD cannot be assured that the plates that passed testing can defeat the threat. DOD also noted several actions DOD and the Army have taken to improve procedures associated with body armor testing. Our responses to these actions are included in comments 2 through 6.

2. The office of the Director of Operational Test and Evaluation's efforts to respond to members of the Armed Services Committees and to address issues raised by the Department of Defense Inspector General were outside the scope of our audit. Therefore, we did not validate the implementation of the actions DOD cited or evaluate their effectiveness in improving test procedures. With regard to the office of the Director of Operational Test and Evaluation's establishing a policy to conduct First Article Testing at government facilities, using a government facility to conduct testing may not necessarily produce improved test results.

3. Regarding the office of the Director of Operational Test and Evaluation's oversight of testing, the office of the Director of Operational Test and Evaluation led the Integrated Product Team and approved the test plans. However, while we were present at the Aberdeen Test Center during Preliminary Design Model testing and First Article Testing, we did not observe on-site monitoring of the testing by the office of the Director of Operational Test and Evaluation staff beyond incidental visits during VIP events and other demonstrations.

4. Regarding the procedures and policies DOD stated were implemented by the Army Test and Evaluation Command to improve testing:

- Only two of the test ranges were completed prior to Preliminary Design Model testing. Two additional test ranges were completed after Preliminary Design Model testing.

- Regarding the certification of the laser scanner measurement device, as noted in our report, the Army had not adequately certified that it was an appropriate tool for body armor testing (see our comment 12).

- The Army's Test Operating Procedure was not completed or implemented until after Preliminary Design Model testing.[1]

- New clay conditioning chambers inside each test range were not constructed until after all testing was completed (see our comment 13).

- The improved velocity measurement accuracy study was not conducted until after all testing was completed.

- Regarding the implementation of electronic data collection and processing for body armor testing, as stated in our report, we observed that not all data are electronically collected. Many types of data are manually collected and are later converted to electronic data storage.

5. Regarding Program Executive Office (PEO) Soldier's efforts to improve the acquisition of personal protection equipment:

- The contract solicitation allowed all prospective body armor manufacturers to compete for new contracts.

- We observed that PEO Soldier did transfer expertise and experience to support Army Acquisition Executive direction that all First Article Testing and lot-acceptance testing be conducted by the Army Test and Evaluation Command.

- The task force that focused on soldier protection was not initiated until February 2009, after all Preliminary Design Model testing and First Article Testing was completed.

- According to Army officials, PEO Soldier instituted a non-destructive test capability that became operational after Preliminary Design Model testing, but prior to First Article Testing.

- PEO Soldier's personal protection evaluation process was described in our previous report—GAO-07-662R. Although we recognized the strength of PEO Soldier's personal protection evaluation process in our earlier report, not all the protections that were in place at that time remain in place. For example, the requirement that testing be

[1] During Preliminary Design Model testing, the most current Army Test Operating Procedure for testing body armor had not been updated since 1975. Test Operations Procedure (TOP), 10-2-506 Ballistic Testing of Personnel Armor Materials. January 6, 1975.

conducted at a National Institute of Justice (NIJ)-certified facility was waived.

6. DOD stated that many of the actions by Army Test and Evaluation Command and PEO Soldier were initiated and improved upon during the course of our review. However, as discussed above, several of these actions were initiated before and during testing, but many of them were not completed until after testing was completed.

7. DOD and the Army stated that Preliminary Design Model testing had achieved its objective to identify those vendor designs that met the performance objectives stated in PEO Soldier's purchase description and that "it is incorrect to state that 'at least two' of the preliminary design models should have failed as they passed in accordance with the modified solicitation." We disagree with these statements. As stated in our report, the most consequential of the deviations from testing protocols we observed involved the measurement of back-face deformation, which did affect final test results. According to original testing protocols, back-face deformation was to be measured *at the deepest point of the depression in the clay backing*. This measure indicates the most force that the armor will allow to be exerted on an individual struck by a bullet. According to Army officials, the deeper the back-face deformation measured in the clay backing, the higher the risk of internal injury or death. DOD and the Army now claim that these solutions passed in accordance with the modified solicitation, which overlooks the fact that the reason the solicitation had to be modified was that Army testers deviated from the testing protocols laid out in the purchase descriptions and did not measure back-face deformation at the deepest point. DOD and the Army also stated in their response that they decided to use the point of aim because they determined it was an accurate and repeatable process. Yet in DOD's detailed comments regarding edge shot locations, DOD acknowledged that there were "potential variances between the actual aim point and impact point during testing." Army Research Laboratory[2] and NIJ-certified laboratories use the benchmark process of measuring back-face deformation at the deepest point, not at the point of aim.

As set forth in our report, at least two solutions passed Preliminary Design Model testing that would have failed if back-face deformation had been measured to the deepest point. This statement came directly from

[2] Member of the Integrated Product Team (IPT).

Aberdeen Test Center officials during a meeting in July 2008, where they specifically told us which two solutions would have failed. We said "at least" two because Army testers did not record deepest point back-face deformation data for the first 30 percent of testing, and therefore there could be more solutions that would have failed had the deepest point been measured during this first portion of the test. Because the Army did not measure back-face deformation to the deepest point, it could not identify whether these two solutions in particular and all the solutions in general met performance requirements. As a result, Army could not waive First Article Testing for successful candidates and was forced to repeat the test to ensure that all solutions did indeed meet requirements. By repeating testing, the Army incurred additional expense and further delays in fielding armor from this solicitation to the soldiers. During the course of our audit, the Army also acknowledged that the Preliminary Design Model testing did not meet its objective because First Article Testing could not be waived without incurring risk to the soldiers.

DOD and the Army stated that, upon discovery of the back-face deformation deviation from the testing protocols described in the purchase descriptions, the Army stopped testing. The Army's Contracting Office was informed of this deviation through a series of questions posed by a vendor who was present at the Vendor Demonstration Day on February 20, 2008. This vendor sent questions to the Contracting Office on February 27 asking whether testers were measuring at the aim point or at the deepest point. This vendor also raised questions about how damage to the soft pack would be recorded and about the location of edge shots. Based on our observations, all of these questions involved issues where Army testers deviated from testing protocols and are discussed in our responses to subsequent comments. The Army did not respond until March 19 and replied that its test procedures complied with solicitation requirements. It was not until Army leadership learned of the vendor's questions and of the deviation in measuring back-face deformation that testing was finally halted on March 27, a full month after the issue came to the Army Test and Evaluation Command's attention.

8. DOD stated that in 2007, prior to the initiation of Preliminary Design Model testing, the Army Test and Evaluation Command, the office of the

Director of Operational Test and Evaluation, and Army leadership all[3] agreed that First Article Testing would be conducted as part of the Army's body armor testing. However, DOD did not provide any documentation dated prior to April 2008—that is, prior to the discovery of the back-face deformation deviation—that suggested that DOD intended to conduct First Article Testing following Preliminary Design Model testing. In July 2008, the Army Test and Evaluation Command and PEO Soldier stated in official written responses to our questions regarding Preliminary Design Model testing that the conduct of First Article Testing became essential following Preliminary Design Model testing because of the Army's measuring back-face deformation at the point of aim as opposed to at the deepest point of deformation. In fact, because of this deviation, DOD could not waive First Article Testing as originally planned and was forced to conduct subsequent tests to verify that the designs that had passed Preliminary Design Model testing met testing requirements. DOD asserted that a multi-phase concept including Preliminary Design Model testing, First Article Testing, and extended ballistic testing to support the development of an improved test standard was briefed to a congressional member and professional staff on November 14, 2007. We were present at this November 14 test overview and strategy/schedule briefing and noted that it did not include plans for First Article Testing to be performed in addition to Preliminary Design Model testing. Excerpts from the slides briefed that day showed Preliminary Design Model (Phase 1) testing and a subsequent ballistic and suitability testing (Phase 2). As indicated in the slides (see fig. 7 and fig. 8) from that November 14 briefing, the Phase 2 test was designed to test the form, fit, and function of those solutions that had passed Preliminary Design Model testing as well as the ballistic statistical confidence tests.[4] According to information we obtained, Phase 2 was never intended to be First Article Testing and was to have no impact on whether or not a solution received a contract.

[3] Omitted from this list of agencies agreeing that First Article Testing was part of the original testing plan are (1) PEO Soldier, the Army's materiel developer and product manager for individual protection equipment being tested and the contracting officer, and (2) the U.S. Army Research, Development, and Engineering Command's Contracting Agency. Both entities told us that First Article Testing was going to be waived.

[4] Form, fit, and function and the test for high ballistic statistical confidence were not part of First Article Testing. Form, fit and function tests involved having soldiers wear the body armor and evaluate its comfort and suitability when performing deployment (war-like) activities—egression from armored vehicles, the double-time run, moving through an obstacle course, and discharging their weapons.

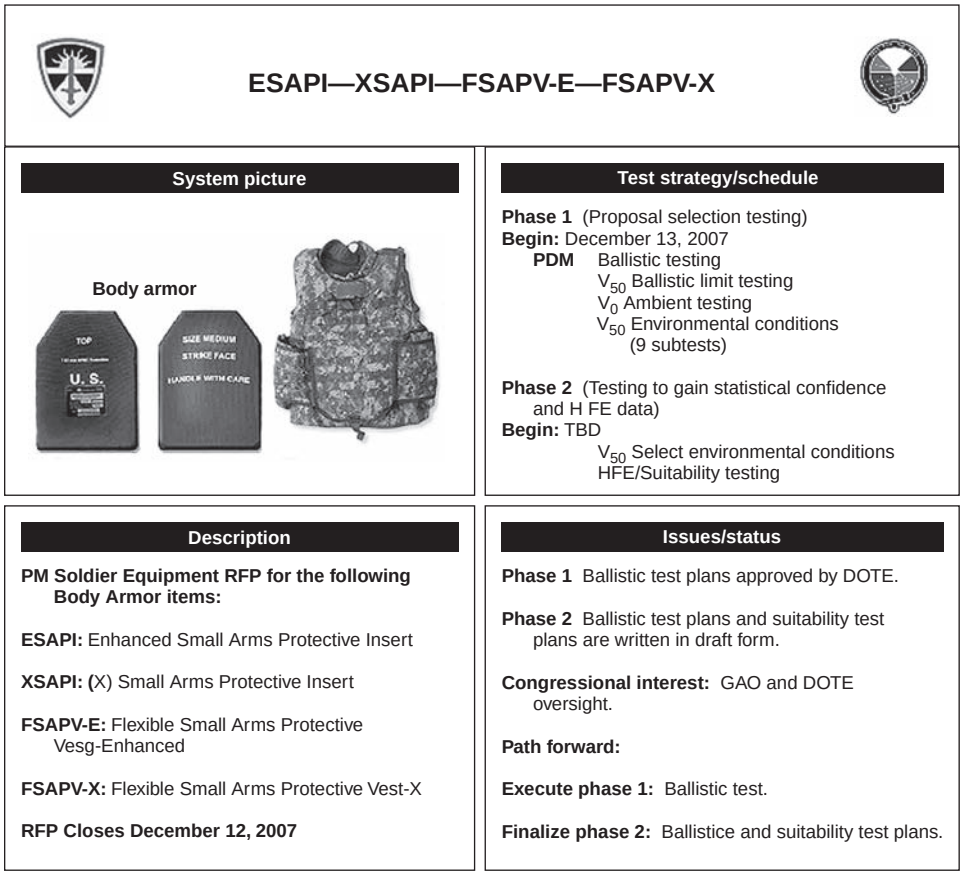
Figure 7: Briefing Slide from DOD's Test Overview (Nov. 14, 2007)

Phase I	1. Physical inspection to ensure compliance with contract requirements and to document material condition.
	2. Ballistic testing and analysis to evaluate test article performance pursuant to contract requirements.
	3. Service test data and all supporting data provided to the Source Selection Technical Factor Chief.
	4. Contract awards(s) made based on best value to the government as determined by the Source Selection Panel analyses of all available data. Bid samples provided by vendors subject to Source Selection Testing IAW First Article Test protocol. Material provided under those awards also provides test articles for Phase 2.
Phase 2	1. Ballistic testing to provide high statistical confidence in material performance.
	2. Suitability testing for form fit, and function. Testing conducted using operarational Soldiers (Ft. Benning, GA to support).
	3. Service T&E report.

**Phase 1 testing described is per the Army test plan approved by CG ATEC on September 11, 2007 and by DOT&E on September 19, 2007.*

Source: Provided by Director of Operational Test and Evaluation at House Armed Services Committee briefing on November 14, 2007.

Figure 8: Briefing Slide from DOD's Test Strategy and Schedule (Nov. 14, 2007)



Source: Provided by Director of Operational Test and Evaluation at House Armed Services Committee Briefing on November 14, 2007. Sensitive ballistic information removed form the "description" box (lower, left-hand-side quadrant).

It was not until after the back-face deformation deviation was discovered that briefing slides and other documentation on test plans and schedules started describing First Article Testing as following Preliminary Design Model testing. For example, as stated by DOD in its comments, the October 2008 briefing to a congressional member and professional staff clearly showed First Article Testing as following Preliminary Design Model testing (Phase 1) and preceding Phase 2. Therefore, it is not clear why DOD's test plan briefings would make no mention of a First Article Testing prior to the back-face deformation measurement deviation while including First Article Testing in subsequent briefings if the plan had always been to conduct both Preliminary Design Model testing and First Article Testing. Furthermore, it is not clear why DOD would intentionally plan at the start of testing to repeat Preliminary Design Model testing (which was supposed

to be performed in accordance with the First Article Testing protocol) with an identical test (First Article Testing) given that it has been the Army's practice to use such Preliminary Design Model testing to meet First Article Testing requirements – a practice that was also supported by the DOD Inspector General and the Army Acquisition Executive after an audit of the Army's body armor testing program.[5] DOD also stated that First Article Testing waivers were not permitted under the body armor solicitation. However, the solicitation and its amendments are unclear as to whether waivers of First Article Testing would be permitted. Nonetheless, in written answers to questions we posed to the Army in July 2008, the Army Test and Evaluation Command and PEO Soldier in a combined response stated that due to the fact that back-face deformation was not measured to the deepest point of penetration during Phase I tests, there would be no waivers of First Article Testing after the contract award.

DOD also stated that it and the Army concluded that First Article Testing had achieved its objective of verifying that contracted vendors could produce, in a full-rate capacity, plates that had passed Preliminary Design Model testing. DOD further stated that it is incorrect to say that First Article Testing did not meet its objective and it is incorrect to assert that three of five vendor designs should have failed First Article Testing. However, our analysis showed that two solutions that passed First Article Testing would have failed if back-face deformations had not been rounded and had been scored as they were during Preliminary Design Model testing.[6] The third solution that passed would have failed if Army testers had correctly scored a shot result as a complete penetration in accordance with the definition of a complete penetration in the purchase description, rather than as a partial penetration. Because questions surround these scoring methods and because DOD and the Army cannot confidently identify whether these vendors can mass produce acceptable plates, we restate that First Article Testing may not have achieved its objective. See comments 12, 10, and 11 regarding DOD's statements about the certification of the laser scanning equipment, the rounding of back-face deformations, and the Aberdeen Test Center's scoring procedures, respectively.

[5] DODIG Report No. D-2008-067, March 31, 2008, DOD Procurement Policy for Body Armor.

[6] Depending on the type of design (i.e., ESAPI or XSAPI) a design can accumulate either 6 or 10 penalty points before being eliminated from consideration. The designs in question were 1.0 point, or one penalty, away from failing.

We agree with DOD that an open dialog with the DOD Inspector General, external test and technology experts, and us will improve the current body armor testing. However, we disagree with DOD's statement that NIJ-certified laboratories lack expertise to provide reliable information on body armor testing issues. Before the current solicitation, the Army relied[7] on these NIJ-certified laboratories for all body armor source selection and lot acceptance tests. The Marine Corps also conducts source selection tests at these facilities. As these independent laboratories have performed numerous tests for the Army conducted in accordance with First Article Testing protocol, we assert that the credentials of these laboratories warrant consideration of their opinions on body armor testing matters.

9. DOD did not concur with our recommendation for an independent evaluation of First Article Testing results before any armor is fielded to soldiers because the First Article Testing achieved its objectives. We disagree with DOD's position that First Article Testing and Preliminary Design Model testing achieved their objectives because we found numerous deviations from testing protocols that allowed solutions to pass testing that otherwise would have failed. Due to these deviations, the majority of which seem to make the testing easier to pass and favor the vendors, we continue to believe that it is necessary to have an independent external expert review the results of First Article Testing and the overall effect of DOD's deviations on those results before the plates are fielded. An independent observer, external to DOD, is best suited to determine the overall impact of DOD's many deviations during the testing associated with this solicitation. Consequently, we have added a matter for Congress to consider directing DOD to either conduct this external review or direct that DOD officially amend its testing protocols to reflect any revised test procedures and repeat First Article Testing.

10. DOD did not concur with our recommendation that the practice of rounding down back-face deformations should be reviewed by external experts because the practice has been used historically by NIJ-certified laboratories. Although DOD acknowledged that the practice of rounding is not adequately described in the testing protocols, it stated that rounding is permitted under American Society for Testing and Materials (ASTM) E-29. The purchase descriptions (attachments 01 and 02 of the solicitation) referenced five ASTM documents, but ASTM E-29 is not referenced and

[7] The law-enforcement community relies on NIJ-certified laboratories to conduct their body armor testing and ensure that their body armor meets law enforcement levels of protection.

therefore is not part of the protocol. The detailed test plans state that solutions shall incur a penalty on deformations greater than 43 millimeters, and the Army is correct that neither the purchase description nor the detailed test plans provide for rounding. During Preliminary Design Model testing, Army testers measured back-face deformations to the hundredths place and did not round. Any deformation between 43.00 and 43.50 received a penalty. During First Article Testing, deformations in this range were rounded down and did not incur a penalty, so the decision to round effectively changed the standard in favor of the vendors. Two solutions passed First Article Testing that would have failed if back-face deformations had been scored without rounding as they were during Preliminary Design Model testing. We recognize that there are other factors, such as the fact that the new laser scanner may overstate back-face deformations that might justify the decision to round down back-face deformations. However, as a stand-alone event, rounding down deformations did change the standard in the middle of the solicitation between Preliminary Design Model testing and First Article Testing. That is why it is important for an independent external expert to review the totality of the test and the Army's deviations from testing protocols to determine the actual effect of this and other deviations.

11. Regarding the incorrect scoring of a complete penetration as a partial penetration, DOD stated that the first layer of soft armor behind the plate serves as a witness plate during testing. If that first layer of soft armor is not penetrated, as determined by the breaking of threads on that first layer of soft armor, the test shot is not scored as a complete penetration in accordance with the PEO Soldier's scoring criteria. However, DOD's position is not consistent with the established testing protocols as evidenced by the following:

(1) We did not observe the use of and the testing protocols do not require the use of a witness plate during testing to determine if a penetration occurred; and

(2) The testing protocols do not state that "the breaking of threads" is the criterion for determining a penetration.

The language of the testing protocols, not undocumented criteria, should be used in scoring and determining penetration results. The criteria for scoring a penetration are found in the current solicitation's protocols. Paragraph 6.6, of each of the purchase descriptions state, under "Definitions: Complete Penetration (CP) for Acceptance Testing--Complete penetrations have occurred when the projectile, fragment of the

projectile, or fragment of the armor material is imbedded or passes into the soft under garment used behind the protective inserts plates" (ESAPIs or XSAPIs).

Our multiple observations and thorough inspection of the soft armor in question revealed that black-grayish particles had penetrated at least three Kevlar layers as evidenced by their frayed, fuzz-like and separated appearance to the naked eye. The black-grayish particles were stopped by the fourth Kevlar layer. DOD acknowledged that figure 6 of our report appears to show evidence of a perforation on the rear of the test plate in question and that the Aberdeen Test Center's subject matter expert found dust particles. These particles are fragments of the projectile or fragments of the armor material that were imbedded and indeed passed into the soft undergarment used behind the protective insert; therefore, the shot should have been ruled a complete penetration according to the testing protocols, increasing the point penalties and causing the design to fail First Article Testing.

DOD's comments stated that we acknowledged there were no broken threads on the first layer of the soft armor. We made no such comment and this consideration is not relevant as the requirement for broken fibers is not consistent with the written testing protocols as we have stated. Of consequence, DOD and Army officials acknowledged that the requirement for broken fibers was not described in the testing protocols or otherwise documented.

In addition to the DOD acknowledgement that an Aberdeen Test Center subject matter expert found particles on the soft body armor, more convincing evidence is the picture of the subject plate. Figure 6 of our report clearly shows the tear in the fibers that were placed behind the plate in question allowing the penetration of the particles found by the Aberdeen Test Center subject matter expert. These particles can only be fragments of the projectile or fragments of the armor material that passed into the soft under garment used behind the protective inserts (plates), confirming our observations of the event and the subsequent incorrect scoring. The shot should have been scored a complete penetration, and the penalty incurred would have caused the design in question to fail First Article Testing.

12. DOD did not concur with our recommendation that the use of the laser scanner needs to be reviewed by experts external to DOD due to the lack of a full evaluation of the scanner's accuracy to measure back-face deformations, to include an evaluation of the software modifications and

operation under actual test conditions. DOD asserted that the laser scanner measurement device provides a superior tool for providing accurate, repeatable, defensible back-face deformation measurements to the deepest point of depression in the clay. We agree that once it is properly certified, tested, and evaluated, the laser may eliminate human errors such as incorrectly selecting the location of the deepest point or piercing the clay with the sharp edge of the caliper and making the depression deeper. However, as we stated, the Army used the laser scanner as a new method to measure back-face deformation without adequately certifying that the scanner could function: (1) in its operational environment, (2) at the required accuracy, (3) in conjunction with its software upgrades, and (4) without overstating deformation measurements.

DOD asserted that the software upgrades did not affect the measurement system of the laser scanner and that these software changes had no effect on the physical measurement process of the back-face deformation measurement that was validated through the certification process. The software upgrades were added after the certification and do include functions[8] to purposely remove spikes and other small crevices on the clay and a smoothing algorithm that changed back-face deformation measurements. We have reviewed these software functions and they do in fact include calculations that change the back-face deformation measurement taken. Furthermore, Army officials told us that additional upgrades to the laser scanner were made after First Article Testing by Aberdeen Test Center to correct a software laser malfunction identified during the subsequent lot acceptance testing of its plates. According to these officials, this previously undetected error caused an overstatement of the back-face deformation measurement taken by several millimeters, calling into question all the measurements taken during First Article Testing. Also, vendors have told us that they have conducted several studies[9] that show that the laser scanner overestimates back-face deformation measurements by about 2 millimeters as compared with measurements taken by digital caliper, thereby over-penalizing vendors'

[8] Software upgrades were not part of the certification process. Some of these software upgrades eliminate the deepest point of depression measurement.

[9] A vendor test showed an approximately 2-millimeter overstatement of back-face deformation measurements by the laser as compared to the caliper.

designs and causing them to fail lot acceptance testing.[10] Furthermore, the laser scanner was certified to an accuracy of 1.0 millimeters, but section 4.9.9.3 of the purchase descriptions requires a device capable of measuring to an accuracy of ±0.1 millimeters. Therefore, the laser does not meet this requirement making the certification invalid. The laser scanner is an unproven measuring device that may reflect a new requirement because the back-face deformation standards are based on measurements obtained with a digital caliper. This raises concerns that results obtained using the laser scanner may be more inconsistent than those obtained using the digital caliper. As we stated in the report, the Aberdeen Test Center has not conducted a side-by-side test of the new laser scanner used during First Article Testing and the digital caliper previously used during Preliminary Design Model testing. Given the discrepancies on back-face deformation measurements we observed and the overstating of the back-face deformation alleged by the vendors, the use of the laser is still called into question. Thus, we continue to support our recommendation that experts independent of DOD review the use of the laser during First Article Testing and that a full evaluation of the laser scanner is imperative to ensure that the tests are repeatable and can be relied upon to certify procurement of armor plates for our military personnel based on results of body armor testing at the Aberdeen Test Center using the laser scanner.

Lastly, DOD stated that the laser scanner is used by the aeronautical industry; however, the Army Test and Evaluation Command officials told us that the scanner had to be customized for testing through various software additions and mounting customizations to mitigate vibrations and other environmental factors. These software additions and customizations change the operation of the scanner.

13. DOD does not concur with our recommendation that experts examine, among other items, "the exposure of clay backing material to rain and other outside environmental conditions as well as the effect of high oven temperatures during storage and conditioning," because it believes that

[10] Lot acceptance testing provides additional ballistic testing that ensures that the plates delivered meet requirements before they are accepted. Two vendors whose designs passed Preliminary Design Model testing and First Article Testing have failed lot acceptance testing and in July 2009 submitted to the Army, in one case, a ruling on a request for equitable adjustment and, in another case, a request to waive contract penalties for late deliveries. These vendors have failed several lot acceptance tests involving tens of thousands of plates that have been rejected by the government because they failed this testing. One vendor is asking for several millions of dollars in payment to compensate for material, labor, and delays as a result of the failed lots.

such conditions had no impact upon First Article Testing results. As detailed in the report, we observed these conditions at different points throughout the testing period. Major variations in materials preparation and testing conditions such as exposure to rain and/or violations of testing protocols merit consideration when analyzing the effectiveness and reliability of First Article Testing.

As one specific example, we described in this report statistically significant differences between the rates of failure in response to one threat on November 13 and the failure rates on all other days of testing but do not use the statistical analysis as the definitive causal explanation for such failure. We observed one major environmental difference in testing conditions that day, the exposure of temperature-conditioned clay to heavy, cold rain in transit to the testing site. After experts confirmed that such variation might be one potential factor relating to overall failure rates on that day, we conducted statistical tests to assess whether failures rates were different on November 13 compared to other dates..

Our assertion that the exposure of the clay to rain may have had an impact on test results is based not solely on our statistical analysis of test results that day; rather, it is also based on our conversations with industry experts, including the clay manufacturer, and on the fact that we witnessed an unusually high number of clay calibration failures during testing that comprised plate designs of multiple vendors, not just the one design that DOD points to as the source for the high failure rate. We observed that the clay conditioning trailer was located approximately 25 feet away from the entrance to the firing lane. The clay blocks, weighing in excess of 200 lbs., were loaded face up onto a cart and then a single individual pulled the cart over approximately 25 feet of gravel to the firing lane entrance. Once there, entry was delayed because the cart had to be positioned just right to get through the firing lane door. Army testers performed all of this without covering the clay[11] to protect it from the rain and the cold, and once inside the clay had significant amounts of water collected on it.

With respect to the unusually high number of clay calibration failures on November 13, there were seven clay calibration drops that were not within

[11]According to Army officials, during subsequent lot acceptance testing tests, Aberdeen Test Center technicians were covering the clay boxes during transport from the conditioning ovens to the lanes.

specifications. Some of these failed clay boxes were discarded in accordance with the testing protocols; however, others were repaired, re-dropped, and used if they had passed the second drop series. These included one plate that was later ruled a no-test and three plates for which the first shot yielded a catastrophic back-face deformation. These were the only three first-shot catastrophic back-face deformations during the whole test, and they all occurred on the same rainy day and involved two different solutions, not just the one that DOD claims performed poorly.

The failure rates of plates as a whole, across all plate designs, were very high this day, and the failures were of both the complete penetration and the back-face deformation variety. Water conducts heat approximately 25 times faster than air, which means the water on the surface cooled the clay considerably faster than the clay would have cooled by air exposure alone. Moreover, Army testers lowered the temperature of the clay conditioning trailers during testing on November 13 and told us that the reason was that the ovens and clay were too hot. This is consistent with what Army subject matter experts and other industry experts told us— that the theoretical effect of having cold rain collecting on hot clay may create a situation where the clay is more susceptible to both complete penetrations because of the colder, harder top layer and to excessive back-face deformations because of the overheated, softer clay beneath the top layer. Finally, the clay manufacturer told us that, although this is an oil-based clay, water can affect the bonding properties of the clay, making it more difficult for wet clay to stick together. This is consistent with what we observed on November 13. After the first shot on one plate, as Army testers were removing the plate from the clay in order to determine the shot result, we observed a large chunk of clay fall to the floor. This clay was simply swept off to the side by the testers. In another instance, as testers were repairing the clay after the calibration drop, one of the testers pulled a long blade over the surface of the clay to smooth it. When he hit the spot where one of the calibration drops had occurred and the clay had been repaired, the blade pulled up the entire divot and the testers had to repair the clay further.

Regarding our use of no-test data, we were strict in the instances where we used this data, see our comment 24.

DOD stated that it was the poor performance of one solution in particular that skewed the results for this day and that this solution failed 70 percent of its shots against Threat D during First Article Testing. DOD's statistic is misleading. This solution failed 100 percent of its shots (6 of 6) on November 13, but only 50 percent for all other test days (7 of 14). Also, the

fact that this solution managed to pass the Preliminary Design Model testing but performed so poorly during First Article Testing raises questions about the repeatability of DOD's and the Army's test practices. Finally, DOD's own analysis confirms that two of the four solutions tested on November 13 performed at their worst level in the test on that day. If the one solution whose plate was questionably ruled a no-test on this day is included in the data, then three of the four solutions performed at their worst level in the test on this day.

DOD said that after testing Aberdeen Test Center completed the planned installation of new clay conditioning chambers inside the test ranges precluding any external environmental conditioning interacting with the clay. We believe it is a step in the right direction that the Aberdeen Test Center has corrected this problem for future testing, but we continue to believe that an external entity needs to evaluate the impact of introducing this new independent variable on this day of First Article Testing.

14. DOD concurred that it should establish a written standard for conducting clay calibration drops but non-concurred that failed blocks were used during testing. DOD asserted that all clay backing material used during testing passed the calibration drop test prior to use. We disagree with this position because the calibration of the clay required by the testing protocols calls for "a series of drops," meaning one series of three drops, not multiple series of three drops as we observed on various occasions. DOD stated that, as a result of our review and the concerns cited in our report, the Aberdeen Test Center established and documented a revised procedure stating that only one repeat of calibration attempt can be made and, if the clay does not pass calibration upon the second attempt, it is reconditioned for later use and a new block of clay is substituted for calibration. Based on the testing protocols, this is still an incorrect procedure to ensure the proper calibration of the clay prior to shooting. The testing protocols do not allow for a repeat series of calibration drops. DOD also says that, upon completion of testing under the current Army solicitation and in coordination with the National Institute of Standards and Technology, the office of the Director of Operational Test and Evaluation and the Army will review the procedures for clay calibration to include repeated calibration attempts and will document any appropriate procedural changes. DOD goes on to say that the NIJ standard as verified by personnel at the National Institute of Standards and Technology does not address specifically the issue of repeating clay calibration tests. However, the Aberdeen Test Center's application of the Army's current solicitation's protocols during testing, and not the NIJ standards, was the subject of our review.

In its comments, DOD acknowledged that the National Institute of Standards and Technology officials recommend only one series of drops for clay calibration, but the Aberdeen Test Center did multiple drops during testing. We are pleased that DOD has agreed to partner with the National Institute of Standards and Technology to conduct experiments to improve the testing community's understanding of clay performance in ballistic testing, but these conversations and studies in our opinion should have occurred prior to testing, not after, as this deviation from testing protocols calls the tests results into question. We reassert that an external entity needs to evaluate the impact of this practice on First Article Testing results.

15. DOD partially concurred with our recommendation and agreed that inconsistencies were identified during testing; however, DOD asserted that the identified inconsistencies did not alter the test results. As stated in our response to DOD's comments on our first recommendation, we do not agree. Our observations clearly show that (1) had the deepest point been used during Preliminary Design Model testing, two designs that passed would have failed and (2) had the Army not rounded First Article Testing results down, two designs that passed would have failed. Further, if the Army had scored the particles (which in their comments to this report DOD acknowledges were imbedded in the shoot pack behind the body armor) according to the testing protocols, a third design that passed First Article Testing would have failed. In all, four out of the five designs that passed Preliminary Design Model testing and First Article Testing would have failed if testing protocols had been followed.

16. DOD partially concurred with our recommendation that, based on the results of the independent expert review of the First Article Testing results, it should evaluate and recertify the accuracy of the laser scanner to the correct standard with all software modifications incorporated and include in this analysis a side-by-side comparison of the laser measurements of the actual back-face deformations with those taken by digital caliper to determine whether laser measurements can meet the standard of the testing protocols. DOD maintains that it performed an independent certification of the laser measurement system and process and that the software changes that occurred did not affect the measurement system in the laser scanner. However, as discussed in comment 12, we do not agree that an adequate, independent certification of the laser measurement system and process was conducted. Based on our observations, we continue to assert that the software changes added after certification did affect the measurement system in the laser.

17. DOD partially concurred with our recommendation for the Secretary of the Army to provide for an independent peer review of the Aberdeen Test Center's body armor testing protocols, facilities, and instrumentation. We agree that a review conducted by a panel of external experts that also includes DOD members could satisfy our recommendation. However, to maintain the independence of this panel, the DOD members should not be composed of personnel from those organizations involved in the body armor testing (such as the office of the Director of Operational Test and Evaluation, the Army Test and Evaluation Command, or PEO Soldier.[12]

18. DOD stated that Aberdeen Test Center had been extensively involved in body armor testing since the 1990s and has performed several tests of body armor plates. We acknowledge that Aberdeen Test Center had conducted limited body armor testing for the initial testing on the Interceptor Body Armor system in the 1990s and have clarified the report to reflect that. However, as acknowledged by DOD, Aberdeen Test Center did not perform any additional testing on that system for PEO Soldier since the 1990s and this lack of experience in conducting source selection testing for that system may have led to the misinterpretations of testing protocols and deviations noted on our report. According to a recent Army Audit Agency report,[13] NIJ testing facilities conducted First Article Testing and lot acceptance testing for the Interceptor Body Armor system prior to this current solicitation. Another reason Aberdeen Test Center could not conduct source selection testing was that in the past Aberdeen Test Center lacked a capability for the production testing of personnel armor systems in a cost-effective manner; the test facilities were old and could not support test requirements for a temperature- and humidity-controlled environment and could not provide enough capacity to support a war-

[12] The Army Test and Evaluation Command performed these tests and the office of the Director of Operational Test and Evaluation provided oversight of Preliminary Design Model testing and First Article Testing for the current solicitation to include determining the scope of testing required and approving the test plans. PEO Soldier provided subject matter experts to advise Army testers, developed the purchase descriptions, and approved test plans. Therefore, these entities are part of the program that needs to be reviewed and are not independent. Additionally, any other individuals and organizations associated with the Preliminary Design Model testing, First Article Testing, or lot acceptance testing should also be excluded.

[13] U.S. Army Audit Agency, Body Armor Testing PEO Soldier; Audit Report: A-2009-0086-ALA, 30 March 2009—Just before the current solicitation, from January 2007 to June 2008, all 27 Army First Article Testing for new designs associated with ESAPIs (four vendors), their associated 1,024 lot acceptance quality assurance ballistic testing, and the long-term environmental conditions testing were all performed in an independent NIJ-certified testing facility.

related workload. The Army has spent about $10 million over the last few years upgrading the existing facilities with state-of-the-art capability to support research and development and production qualification testing for body armor, according to the Army Audit Agency. Army Test and Evaluation Command notes that there were several other tests between 1997 and 2007, but according to Army officials these tests were customer tests not performed in accordance with a First Article Testing protocol. For example, the U.S. Special Operations Command test completed in May 2007 and cited by DOD was a customer test not in accordance with First Article Testing protocol. The Aberdeen Test Center built new lanes and hired and trained contractors to perform the Preliminary Design Model testing and First Article Testing.

19. DOD stated that, to date, it has obligated about $120 million for XSAPI and less than $2 million for ESAPI. However, the value of the 5-year indefinite delivery/indefinite quantity contracts we cited is based on the maximum amount of orders of ESAPI/XSAPI plates that can be purchased under these contracts. Given that the Army has fulfilled the minimum order requirements for this solicitation, the Army could decide to not purchase additional armor based on this solicitation and not incur almost $7.9 billion in costs.

DOD stated in its response that there are only three contracts. However, the Army Contracting Office told us that there were four contracts awarded and provided those contracts to us for our review. Additionally, we witnessed four vendors participating in First Article Testing, all of which had to receive contracts to participate. It is unclear why the Army stated that there were only three contracts.

20. DOD is correct that there is no limit or range specified for the second shot location for the impact subtest. However, this only reinforces that the shot should have been aimed at 1.5 inches, not at 1.0 inch or at various points between 1.0 inch and 1.5 inches. It also does not explain why the Army continued to mark plates as though there were a range for this shot. Army testers would draw lines at approximately 0.75 inches for the inner tolerance and 1.25 inches for the outer tolerance of ESAPI plates. They drew lines at approximately 1.0 inch for the inner tolerance and 1.5 inches for the outer tolerance of XSAPI plates. We measured these lines for every impact test plate and also had Army testers measure some of these lines to

confirm our measurements. We found that of 56 test items,[14] 17 were marked with shot ranges wholly inside of 1.5 inches. The ranges of 30 other test items did include 1.5 inches somewhere in the range, but the center of the range (where Army testers aimed the shot) was still inside of 1.5 inches. Only four test items were marked with ranges centered on 1.5 inches.

DOD may be incorrect in stating that shooting closer to the edge would have increased the risk of a failure for this subtest. For most subtests this may be the case, but according to Army subject matter experts the impact test is different. For the impact test, the plate is dropped onto a concrete surface, striking the crown (center) of the plate. The test is to determine if this weakens the structural integrity of the plate, which could involve various cracks spreading from the center of the plate outward. The reason the requirement for this shot on this subtest is written differently (i.e., to be shot at approximately 1.5 inches from the edge, as opposed to within a range between 0.75 inches and 1.25 inches or between 1.0 inches and 1.5 inches on other subtests) is that it is meant to test the impact's effect on the plate. For this subtest and this shot, there may actually be a higher risk of failure the closer to the center the shot occurs. PEO Soldier representatives acknowledged that the purchase descriptions should have been written more clearly and changed the requirement for this shot to a range of between 1.5 inches and 2.25 inches during First Article Testing. We confirmed that Army testers correctly followed shot location testing protocols during First Article Testing by double-checking the measurements on the firing lane prior to the shooting of the plate. We also note that, although DOD stated the Preliminary Design Model testing shot locations for the impact test complied with the language of the testing protocols, under the revised protocol used during First Article Testing several of these Preliminary Design Model testing impact test shot locations would not have been valid. DOD stated that there was no impact on the outcome of the test, but DOD cannot say that definitively. Because shooting closer to the edge may have favored the vendors in this case, the impact could have been that a solution or solutions may have passed that should not have.

21. The Army stated that "V50 subtests for more robust threats…were executed to the standard protocols." Our observations and analysis of the

[14] Nine test items were not able to be measured due either to the absence of lines or due to damage caused by the impact of the ballistic.

data show that this statement is incorrect. Sections 2.2.3.h(2) of the detailed test plans state:

> "If the first round fired yields a complete penetration, the propellant charge for the second round shall be equal to that of the actual velocity obtained on the first round minus a propellant decrement for 100 ft/s (30 m/s) velocity decrease in order to obtain a partial penetration. If the first round fired yields a partial penetration, the propellant charge for the second round shall be equal to that of the actual velocity obtained on the first round plus a propellant increment for a 50 ft/s (15 m/s) velocity increase in order to obtain a complete penetration. A propellant increment or decrement, as applicable, at 50 ft/s (15 m/s) from actual velocity of last shot shall be used until one partial and one complete penetration is obtained. After obtaining a partial and a complete penetration, the propellant increment or decrement for 50 ft/s (15 m/s) shall be used from the actual velocity of the previous shot."

V50 testing is conducted to discern the velocity at which 50 percent of the shots of a particular threat would penetrate each of the body armor designs. The testing protocols require that, after every shot that is defeated by the body armor, the velocity of the next shot be increased. Whenever a shot penetrates the armor, the velocity should be decreased for the next shot. This increasing and decreasing of the velocities is supposed to be repeated until testers determine the velocity at which 50 percent of the shots will penetrate. In cases in which the armor far exceeds the V50 requirement and is able to defeat the threat for the first six shots, the testing may be halted without discerning the V50 for the plate and the plate may be ruled as passing the requirements. During Preliminary Design Model V50 testing, Army testers would achieve three partial penetrations and then continue to shoot at approximately the same velocity, or lower, for shots 4, 5, and 6 in order to intentionally achieve six partial penetrations. Army testers told us that they did this to conserve plates. According to the testing protocols, Army testers should have continued to increase the charge weight in order to try to achieve a complete penetration and determine a V50 velocity. The effect of this methodology was that solutions were treated inconsistently. Army officials told us that this practice had no effect on which designs passed or failed, which we do not dispute in our report; however, this practice made it impossible to discern the true V50s for these designs based on the results of Preliminary Design Model testing.

22. DOD agreed that Army testers deviated from the testing protocols by measuring back-face deformation at the point of aim. DOD stated that this decision was made by Army leadership in consultation with the office of the Director of Operational Test and Evaluation, because this would not disadvantage any vendor. We agree with DOD that this decision was made by Army leadership in consultation with the office of the Director of Operational Test and Evaluation. We did not independently assess all factors being considered by Army leadership when they made the decision to overrule the Integrated Product Team and the Milestone Decision Authority's initial decision to measure to the deepest point.

DOD also stated that measuring back-face deformation at the point of aim is an accurate and repeatable process. As we pointed out in our previous responses, DOD's own comments regarding DOD's Assertion 3 contradict this statement where DOD writes that there were "potential variances between the actual aim point and impact point during testing." Furthermore, we observed that the aim laser used by Army testers was routinely out of line with where the ballistic was penetrating the yaw card,[15] despite continued adjustments to line up the aim laser with where the ballistic was actually traveling.

DOD stated that it is not possible to know the reference point on a curved object when the deepest deformation point is laterally offset from the aim point. We disagree. DOD acknowledges in its response that PEO Soldier had an internally documented process to account for plate curvature when the deepest point of deformation was laterally offset from the point of aim. The use of correction factor tables is a well-known industry standard that has been in place for years, and this standard practice has been used by NIJ laboratories and is well-known by vendors.

DOD and the Army presented several statistics on the difference between aim point back-face deformation and deepest point back-face deformation in testing and stated that the difference between the two is small. We do not agree with DOD's assertion that a difference of 10.66 millimeters is small. In the case of Preliminary Design Model testing, the difference between measuring at the aim point and at the deepest point was that at least two solutions passed Preliminary Design Model testing that

[15]The yaw card is a piece of paper placed in the intended path of the ballistic and is meant to measure the amount of yaw, or wobble, of the ballistic as it travels through the air. We observed that the hole made by the bullet in the yaw card was routinely not in line with where the aim laser was pointing.

otherwise would have failed. These designs passed subsequent First Article Testing but have gone on to fail lot acceptance testing, raising additional questions regarding the repeatability of the Aberdeen Test Center's testing practices.

DOD asserts that the adoption of the laser scanner measurement technique resolves the problems the Army experienced in measuring back-face deformations completely. We would agree that the laser scanner has the potential to be a useful device but when used in the manner in which Aberdeen Test Center used it – without an adequate certification and without a thorough understanding of how the laser scanner might effectively change the standard for a solution to pass – we do not agree that it resolved back-face deformation measurement issues. Aberdeen Test Center officials told us that they did not know what the accuracy of the laser scanner was as it was used during First Article Testing.

23. DOD acknowledged the shortcoming we identified. DOD then asserted that once the deviation of measuring back-face deformation at the point of aim, rather than at the deepest point of depression was identified, those involved acted decisively to resolve the issue. We disagree based on the timeline of events described in our response to DOD's comments on Preliminary Design Model testing, as well as on the following facts. We were present and observed the Integrated Product Team meeting on March 25 and observed that all members of the Integrated Product Team agreed to start measuring immediately at the deepest point, to score solutions based on this deepest point data, to conserve plates, and then at the end of the testing to make up the tests incorrectly performed during the first third of testing, as needed. We observed Army testers implement this plan the following day. Then, on March 27, Army leadership halted testing for 2 weeks, considered the issue, and then reversed the unanimous decision by the Integrated Product Team and decided to score to the point of aim.

The deviation of scoring solutions based on the back-face deformation at the point of aim created a situation in which the Army could not have confidence in any solution that passed the Preliminary Design Model testing. Because of this, the Army had to repeat testing, in the form of First Article Testing, to determine whether the solutions that had passed Preliminary Design Model testing actually met requirements.

24. DOD did not concur with our finding that rain may have impacted the test results. DOD stated that such conditions had no impact upon First Article Testing results. Our statistical analysis of the test data shows

failure rates to be significantly higher on November 13 than during other days of testing, and our observations taken during that day of testing and our conversations with industry experts familiar with the clay, including the clay manufacturer, suggest the exposure of the clay to the cold, heavy rain on that day may have been the cause of the high failure rates. Our analysis examined the 83 plates tested against the most potent threat, Threat D. The testing protocols required that two shots for the record be taken on each plate. We performed a separate analysis for the 83 first shots taken on these plates from the 83 second shots taken on the plates. These confirmed statistically that the rate of failure on November 13 was significantly higher than the rate of failure on other days. Further, of the 5 plates that experienced first-shot catastrophic failures during testing, 3 of them (60 percent) were tested on November 13 and all 3 of these were due to excessive back-face deformation. Given that only 9 plates were tested on November 13, while 74 were tested during all the other days of testing combined, it is remarkable that 60 percent of all catastrophic failures occurred on that one day of testing.

DOD objected to our inclusion of no-test data in its calculation of first- and second-shot failure rates on November 13. We believe that the inclusion of no-test data is warranted because the Army's exclusion of such plates was made on a post hoc basis after the shots were initially recorded as valid shots and because the rationale for determining the need for a re-test was not always clear. Additionally, we conducted an analysis excluding the no-test plates identified by DOD and that analysis again showed that the failure rate on November 13 was statistically higher than during the other days of testing, even after the exclusions. Excluding the no-test plates, 38 percent of first shots on November 13 (3 of 8) and 88 percent of second shots (7 of 8) failed.

In its response, DOD reports that Aberdeen Test Center's own statistical analysis of test data for Threat D reveals that the observed failure rate on November 13 is attributable to the "poor performance" of one design throughout testing. DOD asserts that its illustration indicates that "Design K was the weakest design on all days with no rain as well as days with rain." DOD's data do not support such a claim. As we have observed, excluding no-test plates, DOD's data are based on 10 tests of two shots each for each of 8 designs (160 cases total). Each shot is treated as an independent trial, an assumption we find tenuous given that a plate's structural integrity might be affected by the first shot. To account for date, DOD subdivides the data into cell sizes far too small to derive reliable statistical inferences about failure rates (between 2 and 6 shots per cell),

as evidenced by the wide confidence intervals illustrated in DOD's visual representation of its analysis.

Among evidence DOD presented to support its claim that Design K was the weakest performing design on both November 13 and other days is failure rate data for four designs that were not tested on the day in question. For two of the three designs tested on November 13 there were only one or two plates tested on November 13, far too few to conduct reliable statistical tests on differences in design performance. For the other type of plate tested on that day (Design L), the three plates tested had a markedly higher failure rate (3 of 6 shots, or 50 percent) on that day than on other days (when it had, in 14 shots, 5 failures, or a 36 percent failure rate). Design K had a failure rate of 6 of 6 shots (100 percent) on the day in question, compared with 8 of 14 shots (57 percent)[16] on other days. Overall, it is impossible to determine from such a small set of tests whether the lack of statistical significance between different designs' failure rates on November 13 and other days results from small sample size or a substantive difference in performance.

Overall, the Army Test and Evaluation Command's design-based analysis cannot distinguish between the potential effects of date and design on failure rates because sufficient comparison data do not exist to conduct the kind of multivariate analysis that might resolve this issue. Because the data alone are inadequate for distinguishing between the potential effects of date and design, we continue to recommend that independent experts evaluate the potential effects of variations in materials preparation and testing conditions, including those occurring on November 13, on overall First Article Testing results.

Additionally, DOD stated that the clay is largely impervious to water. However, as stated in our report, body armor testers from NIJ-certified private laboratories, Army officials experienced in the testing of body armor, body armor manufacturers, and the manufacturer of the clay used told us that getting water on the clay backing material could cause a chemical bonding change on the clay's surface.

[16] According to official test data, only 7 of these 14 shots were failures (50 percent). This is due to the Army's practice of incorrectly rounding down back-face deformations during First Article Testing. One shot that resulted in a back-face deformation of 43.306 was officially rounded down to 43 and not penalized, but had Army testers followed the protocols and not rounded this result down, 8 of the 14 shots would have resulted in penalties.

DOD stated that one of its first actions when bringing in the clay is to scrape the top of the clay to level it. However, this only removes clay that is above the metal edge of the box. Clay that is already at or below the edge of the box is not removed by this scraping. We witnessed several instances in which the blade would remove clay at some points, but leave large portions of the clay surface untouched because the clay was below the edge of the box.

25. See comment 11.

26. The DOD is correct that the one particular example regarding deleting official test data only happened once. Fortunately, the results of the retest were the same as the initial test. After we noted this deficiency, Army officials told us that a new software program was being added that would prevent this from occurring again. DOD also stated that only two persons are authorized and able to modify the laser scanner software. We did not verify this statement; however, we assert that DOD needs to have an auditable trail when any such modifications are made and that it should require supervisory review and documentation or logging of these setting changes.

27. DOD acknowledged that the Army did not formally document significant procedure changes that deviated from established testing protocols or assess the impact of these deviations.

28. In our report we stated that the requirement to test at an NIJ-certified laboratory was withdrawn because the Aberdeen Test Center is not NIJ-certified. DOD's comments on this point do not dispute our statement. Instead, DOD discussed NIJ certification and stated that it does not believe that NIJ certification is appropriate for its test facilities. However, we did not recommend that any DOD test facilities be NIJ-certified or even that NIJ be the outside organization to provide an independent review of the testing practices at Aberdeen Test Center that we did recommend. However, we believe NIJ certification would meet our recommendation for an independent review.

Regarding DOD's comments regarding NIJ certification, DOD asserted that NIJ certification is not appropriate for its test facilities and asserted that there are significant differences between NIJ and U.S. Army body armor test requirements. NIJ certification of a test laboratory and NIJ protocol

for testing personal body armor primarily used by law enforcement officers are two distinct and different issues. Similar to a consumer United Laboratories laboratory certification, an NIJ laboratory certification[17] includes an independent peer review of internal control procedures, management practices, and laboratory practices. This independent peer review is conducted to ensure that there are no conflicts of interest, and that the equipment utilized in the laboratory is safe and reliable. This peer review helps to ensure a reliable, repeatable, and accurate test, regardless of whether the test in question is following a U.S. Army testing protocol or a law enforcement testing protocol. NIJ-certified laboratories have consistently proven to be capable of following an Army testing protocol, which is demonstrated by the fact that NIJ-certified laboratories have conducted previous U.S. Army body armor source selection testing in accordance with First Article Testing protocol, as well as lot acceptance tests. The slide DOD included in its comments is not applicable here because it deals with the difference between testing protocols – the protocols for Army Interceptor Body Armor tests and the NIJ protocol for testing personal body armor primarily used by law enforcement officers. NIJ certification of a laboratory and NIJ certification of body armor for law enforcement purposes are two different things.

29. DOD stated that we were incorrect in asserting that the Army decided to rebuild small arms ballistics testing facilities at Aberdeen Test Center after the 2007 House Armed Services Committee hearing. Instead, DOD stated that the contract to construct additional test ranges at the Aberdeen Test Center Light Armor Range was awarded in September 2006 and that construction was already underway at the time of June 2007 hearing. DOD also stated that this upgrade was not in response to any particular event but was undertaken to meet projected future Army ballistic test requirements. Army officials we spoke with before testing for this solicitation told us that this construction was being completed in order to perform the testing we observed. As of July 2007, the Light Armor Range included two pre-WWII era ballistic lanes and four modern lanes partially completed. However, we noted that, as of July 2007, the lanes we visited were empty and that none of the testing equipment was installed; only the buildings were completed.

[17] The U.S. Department of Justice offers this multi-departmental voluntary compliance program.

In addition to the physical rebuilding of the test sites, the Amy also re-built its workforce to be able to conduct the testing. As stated on page 4 of DOD's comments, PEO Soldier has instituted an effort to transfer testing expertise and experience from PEO Soldier to the Army Test and Evaluation Command. Prior to the start of testing we observed that Aberdeen Test Center hired, transferred in, and contracted for workers to conduct the testing. These workers were then trained by Aberdeen Test Center and conducted pilot tests in order to learn how to conduct body armor testing. We observed parts of this training, in person, and other parts via recorded video. In addition, we spoke with officials during this training and preparation process. From our observations and discussions with Army testers and PEO Soldier officials, we believe this process to have been a restarting of small arms ballistic testing capabilities at Aberdeen Test Center. Based on DOD's comments, we clarified our report to reflect this information.

Appendix III: GAO Contact and Staff Acknowledgments

GAO Contact

William M. Solis, (202) 512-8365

Acknowledgments

In addition to the contact named above, key contributors to this report were Cary Russell, Assistant Director; Michael Aiken; Gary Bianchi; Beverly Breen; Paul Desaulniers; Alfonso Garcia; William Graveline; Mae Jones; Christopher Miller; Anna Maria Ortiz; Danny Owens; Madhav Panwar; Terry Richardson; Michael Shaughnessy; Doug Sloane; Matthew Spiers; Karen Thornton; and John Van Schaik.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday afternoon, GAO posts on its Web site newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to www.gao.gov and select "E-mail Updates."

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's Web site, http://www.gao.gov/ordering.htm.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Ralph Dawn, Managing Director, dawnr@gao.gov, (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548